Classification: Logistic Regression from Data

Machine Learning: Jordan Boyd-Graber
University of Colorado Boulder

LECTURE 3

Slides adapted from William Cohen

Content Questions

Content Questions

Content Questions

Content Questions

Content Questions

Administrivia Questions

Administrivia Questions

Administrivia Questions

Reminder: Logistic Regression

$$P(Y=0|X) = \frac{1}{1+\exp\left[\beta_0 + \sum_i \beta_i X_i\right]} \quad (1)$$

$$P(Y=1|X) = \frac{\exp\left[\beta_0 + \sum_i \beta_i X_i\right]}{1+\exp\left[\beta_0 + \sum_i \beta_i X_i\right]} \quad (2)$$

- Discriminative prediction: $p(y|x)$
- Classification uses: ad placement, spam detection
- What we didn't talk about is how to learn β from data

Logistic Regression: Objective Function

$$\mathcal{L} \equiv \ln p(Y \mid X, \beta) = \sum_j \ln p(y^{(j)} \mid x^{(j)}, \beta) \quad (3)$$

$$= \sum_j y^{(j)} \left(\beta_0 + \sum_i \beta_i x_i^{(j)} \right) - \ln \left[1 + \exp \left(\beta_0 + \sum_i \beta_i x_i^{(j)} \right) \right] \quad (4)$$

Algorithm

1. Initialize a vector B to be all zeros
2. For $t = 1, \dots, T$
   - For each example $\vec{x}_i, y_i$ and feature j:
     - Compute $\pi_i \equiv \Pr(y_i = 1 \mid \vec{x}_i)$
     - Set $\beta[j] = \beta[j]' + \lambda(y_i - \pi_i)x_i$
3. Output the parameters $\beta_1, \dots, \beta_d$.

Example Documents

$$\beta[j] = \beta[j] + \lambda(y_i - \pi_i)x_i$$

$$\vec{\beta} = \langle \beta_{bias} = 0, \beta_A = 0, \beta_B = 0, \beta_C = 0, \beta_D = 0 \rangle$$

$y_1 = \mathbf{1}$

A A A A B B B C

$y_2 = \mathbf{0}$

B C C C D D D D

(Assume step size $\lambda = 1.0$.)

You first see the positive example. First, compute π_1

Example Documents

$$\beta[j] = \beta[j] + \lambda(y_i - \pi_i)x_i$$
$$\vec{\beta} = \langle 0, 0, 0, 0, 0 \rangle$$

$y_1 = \mathbf{1}$

A A A A B B B C

$y_2 = \mathbf{0}$

B C C C D D D D

(Assume step size $\lambda = 1.0$.)

You first see the positive example. First, compute π_1

$\pi_1 = \Pr(y_1 = 1 \mid \vec{x_1}) = \frac{\exp \beta^T x_i}{1 + \exp \beta^T x_i} =$

Example Documents

$$\beta[j] = \beta[j] + \lambda(y_i - \pi_i)x_i$$
$$\vec{\beta} = \langle 0, 0, 0, 0, 0 \rangle$$

$y_1 = \mathbf{1}$

A A A A B B B C

$y_2 = \mathbf{0}$

B C C C D D D D

(Assume step size $\lambda = 1.0$.)

You first see the positive example. First, compute π_1

$\pi_1 = \Pr(y_1 = 1 \mid \vec{x_1}) = \frac{\exp \beta^T x_i}{1 + \exp \beta^T x_i} = \frac{\exp 0}{\exp 0 + 1} = 0.5$

Example Documents

$$\beta[j] = \beta[j] + \lambda(y_i - \pi_i)x_i$$
$$\vec{\beta} = \langle 0, 0, 0, 0, 0 \rangle$$

$y_1 = \mathbf{1}$	$y_2 = \mathbf{0}$
A A A A B B B C	B C C C D D D D

(Assume step size $\lambda = 1.0$.)

$\pi_1 = 0.5$ What's the update for β_{bias}?

Example Documents

$$\beta[j] = \beta[j] + \lambda(y_i - \pi_i)x_i$$
$$\vec{\beta} = \langle 0, 0, 0, 0, 0 \rangle$$

$y_1 = \mathbf{1}$

A A A A B B B C

$y_2 = \mathbf{0}$

B C C C D D D D

(Assume step size $\lambda = 1.0$.)

What's the update for β_{bias}?

$$\beta_{bias} = \beta'_{bias} + \lambda \cdot (y_1 - \pi_1) \cdot x_{1,bias} = 0.0 + 1.0 \cdot (1.0 - 0.5) \cdot 1.0$$

Example Documents

$$\beta[j] = \beta[j] + \lambda(y_i - \pi_i)x_i$$
$$\vec{\beta} = \langle 0,0,0,0,0 \rangle$$

$y_1 = \mathbf{1}$

A A A A B B B C

$y_2 = \mathbf{0}$

B C C C D D D D

(Assume step size $\lambda = 1.0$.)

What's the update for β_{bias}?

$\beta_{bias} = \beta'_{bias} + \lambda \cdot (y_1 - \pi_1) \cdot x_{1,bias} = 0.0 + 1.0 \cdot (1.0 - 0.5) \cdot 1.0$ =0.5

Example Documents

$$\beta[j] = \beta[j] + \lambda(y_i - \pi_i)x_i$$
$$\vec{\beta} = \langle 0, 0, 0, 0, 0 \rangle$$

$y_1 = \mathbf{1}$	$y_2 = \mathbf{0}$
A A A A B B B C	B C C C D D D D

(Assume step size $\lambda = 1.0$.)

What's the update for β_A?

Example Documents

$$\beta[j] = \beta[j] + \lambda(y_i - \pi_i)x_i$$
$$\vec{\beta} = \langle 0, 0, 0, 0, 0 \rangle$$

$y_1 = \mathbf{1}$

A A A A B B B C

$y_2 = \mathbf{0}$

B C C C D D D D

(Assume step size $\lambda = 1.0$.)

What's the update for β_A?

$\beta_A = \beta'_A + \lambda \cdot (y_1 - \pi_1) \cdot x_{1,A} = 0.0 + 1.0 \cdot (1.0 - 0.5) \cdot 4.0$

Example Documents

$$\beta[j] = \beta[j] + \lambda(y_i - \pi_i)x_i$$
$$\vec{\beta} = \langle 0, 0, 0, 0, 0 \rangle$$

$y_1 = \mathbf{1}$

A A A A B B B C

$y_2 = \mathbf{0}$

B C C C D D D D

(Assume step size $\lambda = 1.0$.)

What's the update for β_A?

$\beta_A = \beta'_A + \lambda \cdot (y_1 - \pi_1) \cdot x_{1,A} = 0.0 + 1.0 \cdot (1.0 - 0.5) \cdot 4.0$ =2.0

Example Documents

$$\beta[j] = \beta[j] + \lambda(y_i - \pi_i)x_i$$
$$\vec{\beta} = \langle 0, 0, 0, 0, 0 \rangle$$

$y_1 = \mathbf{1}$

A A A A B B B C

$y_2 = \mathbf{0}$

B C C C D D D D

(Assume step size $\lambda = 1.0$.)

What's the update for β_B?

Example Documents

$$\beta[j] = \beta[j] + \lambda(y_i - \pi_i)x_i$$

$$\vec{\beta} = \langle 0, 0, 0, 0, 0 \rangle$$

$y_1 = \mathbf{1}$	$y_2 = \mathbf{0}$
A A A A B B B C	B C C C D D D D

(Assume step size $\lambda = 1.0$.)

What's the update for β_B?

$\beta_B = \beta'_B + \lambda \cdot (y_1 - \pi_1) \cdot x_{1,B} = 0.0 + 1.0 \cdot (1.0 - 0.5) \cdot 3.0$

Example Documents

$$\beta[j] = \beta[j] + \lambda(y_i - \pi_i)x_i$$
$$\vec{\beta} = \langle 0, 0, 0, 0, 0 \rangle$$

$y_1 = \mathbf{1}$	$y_2 = \mathbf{0}$
A A A A B B B C	B C C C D D D D

(Assume step size $\lambda = 1.0$.)

What's the update for β_B?

$\beta_B = \beta'_B + \lambda \cdot (y_1 - \pi_1) \cdot x_{1,B} = 0.0 + 1.0 \cdot (1.0 - 0.5) \cdot 3.0$ =1.5

Example Documents

$$\beta[j] = \beta[j] + \lambda(y_i - \pi_i)x_i$$

$$\vec{\beta} = \langle 0, 0, 0, 0, 0 \rangle$$

$y_1 = \mathbf{1}$

A A A A B B B C

$y_2 = \mathbf{0}$

B C C C D D D D

(Assume step size $\lambda = 1.0$.)

What's the update for β_C?

Example Documents

$$\beta[j] = \beta[j] + \lambda(y_i - \pi_i)x_i$$
$$\vec{\beta} = \langle 0,0,0,0,0 \rangle$$

$y_1 = \mathbf{1}$	$y_2 = \mathbf{0}$
A A A A B B B C	B C C C D D D D

(Assume step size $\lambda = 1.0$.)

What's the update for β_C?

$\beta_C = \beta'_C + \lambda \cdot (y_1 - \pi_1) \cdot x_{1,C} = 0.0 + 1.0 \cdot (1.0 - 0.5) \cdot 1.0$

Example Documents

$$\beta[j] = \beta[j] + \lambda(y_i - \pi_i)x_i$$
$$\vec{\beta} = \langle 0, 0, 0, 0, 0 \rangle$$

$y_1 = \mathbf{1}$

A A A A B B B C

$y_2 = \mathbf{0}$

B C C C D D D D

(Assume step size $\lambda = 1.0$.)

What's the update for β_C?

$\beta_C = \beta'_C + \lambda \cdot (y_1 - \pi_1) \cdot x_{1,C} = 0.0 + 1.0 \cdot (1.0 - 0.5) \cdot 1.0$ =0.5

Example Documents

$$\beta[j] = \beta[j] + \lambda(y_i - \pi_i)x_i$$
$$\vec{\beta} = \langle 0, 0, 0, 0, 0 \rangle$$

$y_1 = \mathbf{1}$

A A A A B B B C

$y_2 = \mathbf{0}$

B C C C D D D D

(Assume step size $\lambda = 1.0$.)

What's the update for β_D?

Example Documents

$$\beta[j] = \beta[j] + \lambda(y_i - \pi_i)x_i$$
$$\vec{\beta} = \langle 0, 0, 0, 0, 0 \rangle$$

$y_1 = \mathbf{1}$

A A A A B B B C

$y_2 = \mathbf{0}$

B C C C D D D D

(Assume step size $\lambda = 1.0$.)

What's the update for β_D?

$\beta_D = \beta'_D + \lambda \cdot (y_1 - \pi_1) \cdot x_{1,D} = 0.0 + 1.0 \cdot (1.0 - 0.5) \cdot 0.0$

Example Documents

$$\beta[j] = \beta[j] + \lambda(y_i - \pi_i)x_i$$
$$\vec{\beta} = \langle 0, 0, 0, 0, 0 \rangle$$

$y_1 = \mathbf{1}$	$y_2 = \mathbf{0}$
A A A A B B B C	B C C C D D D D

(Assume step size $\lambda = 1.0$.)

What's the update for β_D?

$\beta_D = \beta'_D + \lambda \cdot (y_1 - \pi_1) \cdot x_{1,D} = 0.0 + 1.0 \cdot (1.0 - 0.5) \cdot 0.0$ =0.0

Example Documents

$$\beta[j] = \beta[j] + \lambda(y_i - \pi_i)x_i$$

$$\vec{\beta} = \langle .5, 2, 1.5, 0.5, 0 \rangle$$

$y_1 = \mathbf{1}$	$y_2 = \mathbf{0}$
A A A A B B B C	B C C C D D D D

(Assume step size $\lambda = 1.0$.)

Now you see the negative example. What's π_2?

Example Documents

$$\beta[j] = \beta[j] + \lambda(y_i - \pi_i)x_i$$

$$\vec{\beta} = \langle .5, 2, 1.5, 0.5, 0 \rangle$$

$y_1 = \mathbf{1}$	$y_2 = \mathbf{0}$
A A A A B B B C	B C C C D D D D

(Assume step size $\lambda = 1.0$.)

Now you see the negative example. What's π_2?

$$\pi_2 = \Pr(y_2 = 1 \mid \vec{x_2}) = \frac{\exp \beta^T x_i}{1 + \exp \beta^T x_i} = \frac{\exp[.5 + 1.5 + 1.5 + 0]}{\exp[.5 + 1.5 + 1.5 + 0] + 1} =$$

Example Documents

$$\beta[j] = \beta[j] + \lambda(y_i - \pi_i)x_i$$

$$\vec{\beta} = \langle .5, 2, 1.5, 0.5, 0 \rangle$$

$y_1 = \mathbf{1}$

A A A A B B B C

$y_2 = \mathbf{0}$

B C C C D D D D

(Assume step size $\lambda = 1.0$.)

Now you see the negative example. What's π_2?

$$\pi_2 = \Pr(y_2 = 1 \mid \vec{x_2}) = \frac{\exp \beta^T x_i}{1 + \exp \beta^T x_i} = \frac{\exp[.5 + 1.5 + 1.5 + 0]}{\exp[.5 + 1.5 + 1.5 + 0] + 1} = 0.97$$

Example Documents

$$\beta[j] = \beta[j] + \lambda(y_i - \pi_i)x_i$$
$$\vec{\beta} = \langle .5, 2, 1.5, 0.5, 0 \rangle$$

$y_1 = \mathbf{1}$	$y_2 = \mathbf{0}$
A A A A B B B C	B C C C D D D D

(Assume step size $\lambda = 1.0$.)

Now you see the negative example. What's π_2?

$\pi_2 = 0.97$

What's the update for β_{bias}?

Example Documents

$$\beta[j] = \beta[j] + \lambda(y_i - \pi_i)x_i$$
$$\vec{\beta} = \langle .5, 2, 1.5, 0.5, 0 \rangle$$

$y_1 = \mathbf{1}$

A A A A B B B C

$y_2 = \mathbf{0}$

B C C C D D D D

(Assume step size $\lambda = 1.0$.)

What's the update for β_{bias}?

$\beta_{bias} = \beta'_{bias} + \lambda \cdot (y_2 - \pi_2) \cdot x_{2,bias} = 0.5 + 1.0 \cdot (0.0 - 0.97) \cdot 1.0$

Example Documents

$$\beta[j] = \beta[j] + \lambda(y_i - \pi_i)x_i$$
$$\vec{\beta} = \langle .5, 2, 1.5, 0.5, 0 \rangle$$

$y_1 = \mathbf{1}$	$y_2 = \mathbf{0}$
A A A A B B B C	B C C C D D D D

(Assume step size $\lambda = 1.0$.)

What's the update for β_{bias}?

$\beta_{bias} = \beta'_{bias} + \lambda \cdot (y_2 - \pi_2) \cdot x_{2,bias} = 0.5 + 1.0 \cdot (0.0 - 0.97) \cdot 1.0$ =-0.47

Example Documents

$$\beta[j] = \beta[j] + \lambda(y_i - \pi_i)x_i$$

$$\vec{\beta} = \langle .5, 2, 1.5, 0.5, 0 \rangle$$

$y_1 = \mathbf{1}$	$y_2 = \mathbf{0}$
A A A A B B B C	B C C C D D D D

(Assume step size $\lambda = 1.0$.)

What's the update for β_A?

Example Documents

$$\beta[j] = \beta[j] + \lambda(y_i - \pi_i)x_i$$
$$\vec{\beta} = \langle .5, 2, 1.5, 0.5, 0 \rangle$$

$y_1 = \mathbf{1}$	$y_2 = \mathbf{0}$
A A A A B B B C	B C C C D D D D

(Assume step size $\lambda = 1.0$.)

What's the update for β_A?

$\beta_A = \beta'_A + \lambda \cdot (y_2 - \pi_2) \cdot x_{2,A} = 2.0 + 1.0 \cdot (0.0 - 0.97) \cdot 0.0$

Example Documents

$$\beta[j] = \beta[j] + \lambda(y_i - \pi_i)x_i$$
$$\vec{\beta} = \langle .5, 2, 1.5, 0.5, 0 \rangle$$

$y_1 = \mathbf{1}$	$y_2 = \mathbf{0}$
A A A A B B B C	B C C C D D D D

(Assume step size $\lambda = 1.0$.)

What's the update for β_A?

$\beta_A = \beta'_A + \lambda \cdot (y_2 - \pi_2) \cdot x_{2,A} = 2.0 + 1.0 \cdot (0.0 - 0.97) \cdot 0.0 \quad = 2.0$

Example Documents

$$\beta[j] = \beta[j] + \lambda(y_i - \pi_i)x_i$$
$$\vec{\beta} = \langle .5, 2, 1.5, 0.5, 0 \rangle$$

$y_1 = \mathbf{1}$	$y_2 = \mathbf{0}$
A A A A B B B C	B C C C D D D D

(Assume step size $\lambda = 1.0$.)

What's the update for β_B?

Example Documents

$$\beta[j] = \beta[j] + \lambda(y_i - \pi_i)x_i$$
$$\vec{\beta} = \langle .5, 2, 1.5, 0.5, 0 \rangle$$

$y_1 = \mathbf{1}$

A A A A B B B C

$y_2 = \mathbf{0}$

B C C C D D D D

(Assume step size $\lambda = 1.0$.)

What's the update for β_B?

$\beta_B = \beta'_B + \lambda \cdot (y_2 - \pi_2) \cdot x_{2,B} = 1.5 + 1.0 \cdot (0.0 - 0.97) \cdot 1.0$

Example Documents

$$\beta[j] = \beta[j] + \lambda(y_i - \pi_i)x_i$$

$$\vec{\beta} = \langle .5, 2, 1.5, 0.5, 0 \rangle$$

$y_1 = \mathbf{1}$	$y_2 = \mathbf{0}$
A A A A B B B C	B C C C D D D D

(Assume step size $\lambda = 1.0$.)

What's the update for β_B?

$\beta_B = \beta_B' + \lambda \cdot (y_2 - \pi_2) \cdot x_{2,B} = 1.5 + 1.0 \cdot (0.0 - 0.97) \cdot 1.0$ =0.53

Example Documents

$$\beta[j] = \beta[j] + \lambda(y_i - \pi_i)x_i$$
$$\vec{\beta} = \langle .5, 2, 1.5, 0.5, 0 \rangle$$

$y_1 = \mathbf{1}$

A A A A B B B C

$y_2 = \mathbf{0}$

B C C C D D D D

(Assume step size $\lambda = 1.0$.)

What's the update for β_C?

Example Documents

$$\beta[j] = \beta[j] + \lambda(y_i - \pi_i)x_i$$

$$\vec{\beta} = \langle .5, 2, 1.5, 0.5, 0 \rangle$$

$y_1 = \mathbf{1}$

A A A A B B B C

$y_2 = \mathbf{0}$

B C C C D D D D

(Assume step size $\lambda = 1.0$.)

What's the update for β_C?

$\beta_C = \beta'_C + \lambda \cdot (y_2 - \pi_2) \cdot x_{2,C} = 0.5 + 1.0 \cdot (0.0 - 0.97) \cdot 3.0$

Example Documents

$$\beta[j] = \beta[j] + \lambda(y_i - \pi_i)x_i$$
$$\vec{\beta} = \langle .5, 2, 1.5, 0.5, 0 \rangle$$

$y_1 = \mathbf{1}$	$y_2 = \mathbf{0}$
A A A A B B B C	B C C C D D D D

(Assume step size $\lambda = 1.0$.)

What's the update for β_C?

$\beta_C = \beta'_C + \lambda \cdot (y_2 - \pi_2) \cdot x_{2,C} = 0.5 + 1.0 \cdot (0.0 - 0.97) \cdot 3.0$ =-2.41

Example Documents

$$\beta[j] = \beta[j] + \lambda(y_i - \pi_i)x_i$$
$$\vec{\beta} = \langle .5, 2, 1.5, 0.5, 0 \rangle$$

$y_1 = \mathbf{1}$	$y_2 = \mathbf{0}$
A A A A B B B C	B C C C D D D D

(Assume step size $\lambda = 1.0$.)

What's the update for β_D?

Example Documents

$$\beta[j] = \beta[j] + \lambda(y_i - \pi_i)x_i$$
$$\vec{\beta} = \langle .5, 2, 1.5, 0.5, 0 \rangle$$

$y_1 = \mathbf{1}$

A A A A B B B C

$y_2 = \mathbf{0}$

B C C C D D D D

(Assume step size $\lambda = 1.0$.)

What's the update for β_D?

$\beta_D = \beta'_D + \lambda \cdot (y_2 - \pi_2) \cdot x_{2,D} = 0.0 + 1.0 \cdot (0.0 - 0.97) \cdot 4.0$

Example Documents

$$\beta[j] = \beta[j] + \lambda(y_i - \pi_i)x_i$$
$$\vec{\beta} = \langle .5, 2, 1.5, 0.5, 0 \rangle$$

$y_1 = \mathbf{1}$	$y_2 = \mathbf{0}$
A A A A B B B C	B C C C D D D D

(Assume step size $\lambda = 1.0$.)

What's the update for β_D?

$\beta_D = \beta'_D + \lambda \cdot (y_2 - \pi_2) \cdot x_{2,D} = 0.0 + 1.0 \cdot (0.0 - 0.97) \cdot 4.0$ =-3.88

How does the gradient change with regularization?

- You can do your normal update

How does the gradient change with regularization?

- You can do your normal update
- Then

$$\beta_j = \beta_j' - \lambda 2 \mu \beta_j = \beta_j' \cdot (1 - 2\lambda\mu) \tag{5}$$

How does the gradient change with regularization?

- You can do your normal update
- Then

$$\beta_j = \beta_j' - \lambda 2\mu\beta_j = \beta_j' \cdot (1 - 2\lambda\mu) \tag{5}$$

- Doesn't depend on X or Y. Just makes all your weights smaller

How does the gradient change with regularization?

- You can do your normal update
- Then

$$\beta_j = \beta_j' - \lambda 2 \mu \beta_j = \beta_j' \cdot (1 - 2\lambda\mu) \tag{5}$$

- Doesn't depend on X or Y. Just makes all your weights smaller
- But difficult to update every feature every time (if there are many features)

How does the gradient change with regularization?

- You can do your normal update
- Then

$$\beta_j = \beta_j' - \lambda 2\mu\beta_j = \beta_j' \cdot (1 - 2\lambda\mu) \tag{5}$$

- Doesn't depend on X or Y. Just makes all your weights smaller
- But difficult to update every feature every time (if there are many features)
- Following this up, we note that we can perform m successive "regularization" updates by letting $\beta_j = \beta_j' \cdot (1 - 2\lambda\mu)^{m_j}$

Basic Idea

Don't perform regularization updates for zero-valued x_j's, but instead to simply keep track of how many such updates would need to be performed to update β_j

Example Documents (Regularized)

$$\beta[j] = (\beta[j]' + \lambda(y-p)x_i) \cdot (1-2\lambda\mu)^{m_j}$$
$$\vec{\beta} = \langle 0,0,0,0,0 \rangle$$

$y_1 = 1$

A A A A B B B C

$y_2 = 0$

B C C C D D D D

Assume step size $\lambda = 1.0$ and $\mu = \frac{1}{4}$.

Example Documents (Regularized)

$$\beta[j] = (\beta[j]' + \lambda(y-p)x_i) \cdot (1-2\lambda\mu)^{m_j}$$
$$\vec{\beta} = \langle 0,0,0,0,0 \rangle$$

$y_1 = 1$

A A A A B B B C

$y_2 = 0$

B C C C D D D D

Assume step size $\lambda = 1.0$ and $\mu = \frac{1}{4}$.

You first see the positive example. π_1 is still 0.5.

Example Documents (Regularized)

$$\beta[j] = (\beta[j]' + \lambda(y - p)x_i) \cdot (1 - 2\lambda\mu)^{m_j}$$
$$\vec{\beta} = \langle 0, 0, 0, 0, 0 \rangle$$

$y_1 = 1$

A A A A B B B C

$y_2 = 0$

B C C C D D D D

Assume step size $\lambda = 1.0$ and $\mu = \frac{1}{4}$.

You first see the positive example. π_1 is still 0.5.
What's the update for β_{bias}?

Example Documents (Regularized)

$$\beta[j] = (\beta[j]' + \lambda(y - p)x_i) \cdot (1 - 2\lambda\mu)^{m_j}$$
$$\vec{\beta} = \langle 0, 0, 0, 0, 0 \rangle$$

$y_1 = 1$	$y_2 = 0$
A A A A B B B C	B C C C D D D D

Assume step size $\lambda = 1.0$ and $\mu = \frac{1}{4}$.

What's the update for β_{bias}?

$\beta_{bias} = \left(\beta'_{bias} + \lambda \cdot (y_1 - \pi_1) \cdot x_{1,bias}\right)(1 - 2 \cdot \lambda \cdot \mu)^{m_{bias}} =$
$(0.0 + 1.0 \cdot (1.0 - 0.5) \cdot 1.0)\left(1 - 2 \cdot 1.0 \cdot \frac{1}{4}\right)^1$

Example Documents (Regularized)

$$\beta[j] = (\beta[j]' + \lambda(y - p)x_i) \cdot (1 - 2\lambda\mu)^{m_j}$$
$$\vec{\beta} = \langle 0, 0, 0, 0, 0 \rangle$$

$y_1 = 1$	$y_2 = 0$
A A A A B B B C	B C C C D D D D

Assume step size $\lambda = 1.0$ and $\mu = \frac{1}{4}$.

What's the update for β_{bias}?

$\beta_{bias} = \left(\beta'_{bias} + \lambda \cdot (y_1 - \pi_1) \cdot x_{1,bias}\right) (1 - 2 \cdot \lambda \cdot \mu)^{m_{bias}} =$

$(0.0 + 1.0 \cdot (1.0 - 0.5) \cdot 1.0) \left(1 - 2 \cdot 1.0 \cdot \frac{1}{4}\right)^1 \quad = 0.25$

Example Documents (Regularized)

$$\beta[j] = (\beta[j]' + \lambda(y-p)x_i) \cdot (1 - 2\lambda\mu)^{m_j}$$
$$\vec{\beta} = \langle 0,0,0,0,0 \rangle$$

$y_1 = 1$

A A A A B B B C

$y_2 = 0$

B C C C D D D D

Assume step size $\lambda = 1.0$ and $\mu = \frac{1}{4}$.

What's the update for β_A?

Example Documents (Regularized)

$$\beta[j] = (\beta[j]' + \lambda(y-p)x_i) \cdot (1-2\lambda\mu)^{m_j}$$
$$\vec{\beta} = \langle 0,0,0,0,0 \rangle$$

$y_1 = 1$	$y_2 = 0$
A A A A B B B C	B C C C D D D D

Assume step size $\lambda = 1.0$ and $\mu = \frac{1}{4}$.

What's the update for β_A?

$\beta_A = \left(\beta_A' + \lambda \cdot (y_1 - \pi_1) \cdot x_{1,A}\right)(1 - 2 \cdot \lambda \cdot \mu)^{m_A} =$

$(0.0 + 1.0 \cdot (1.0 - 0.5) \cdot 4.0)\left(1 - 2 \cdot 1.0 \cdot \frac{1}{4}\right)^1$

Example Documents (Regularized)

$$\beta[j] = (\beta[j]' + \lambda(y-p)x_i) \cdot (1-2\lambda\mu)^{m_j}$$
$$\vec{\beta} = \langle 0, 0, 0, 0, 0 \rangle$$

$y_1 = 1$

A A A A B B B C

$y_2 = 0$

B C C C D D D D

Assume step size $\lambda = 1.0$ and $\mu = \frac{1}{4}$.

What's the update for β_A?

$\beta_A = \left(\beta'_A + \lambda \cdot (y_1 - \pi_1) \cdot x_{1,A}\right)(1 - 2 \cdot \lambda \cdot \mu)^{m_A} =$

$(0.0 + 1.0 \cdot (1.0 - 0.5) \cdot 4.0)\left(1 - 2 \cdot 1.0 \cdot \frac{1}{4}\right)^1$ $=1.0$

Example Documents (Regularized)

$$\beta[j] = (\beta[j]' + \lambda(y-p)x_i) \cdot (1-2\lambda\mu)^{m_j}$$
$$\vec{\beta} = \langle 0, 0, 0, 0, 0 \rangle$$

$y_1 = 1$	$y_2 = 0$
A A A A B B B C	B C C C D D D D

Assume step size $\lambda = 1.0$ and $\mu = \frac{1}{4}$.

What's the update for β_B?

Example Documents (Regularized)

$$\beta[j] = (\beta[j]' + \lambda(y-p)x_i) \cdot (1-2\lambda\mu)^{m_j}$$
$$\vec{\beta} = \langle 0,0,0,0,0 \rangle$$

$y_1 = 1$	$y_2 = 0$
A A A A B B B C	B C C C D D D D

Assume step size $\lambda = 1.0$ and $\mu = \frac{1}{4}$.

What's the update for β_B?

$$\beta_B = \left(\beta_B' + \lambda \cdot (y_1 - \pi_1) \cdot x_{1,B}\right)(1 - 2 \cdot \lambda \cdot \mu)^{m_B} =$$
$$(0.0 + 1.0 \cdot (1.0 - 0.5) \cdot 3.0)\left(1 - 2 \cdot 1.0 \cdot \frac{1}{4}\right)^1$$

Example Documents (Regularized)

$$\beta[j] = (\beta[j]' + \lambda(y - p)x_i) \cdot (1 - 2\lambda\mu)^{m_j}$$
$$\vec{\beta} = \langle 0, 0, 0, 0, 0 \rangle$$

$y_1 = 1$	$y_2 = 0$
A A A A B B B C	B C C C D D D D

Assume step size $\lambda = 1.0$ and $\mu = \frac{1}{4}$.

What's the update for β_B?

$\beta_B = \left(\beta_B' + \lambda \cdot (y_1 - \pi_1) \cdot x_{1,B}\right)(1 - 2 \cdot \lambda \cdot \mu)^{m_B} =$

$(0.0 + 1.0 \cdot (1.0 - 0.5) \cdot 3.0)\left(1 - 2 \cdot 1.0 \cdot \frac{1}{4}\right)^1 \quad = 0.75$

Example Documents (Regularized)

$$\beta[j] = (\beta[j]' + \lambda(y - p)x_i) \cdot (1 - 2\lambda\mu)^{m_j}$$
$$\vec{\beta} = \langle 0, 0, 0, 0, 0 \rangle$$

$y_1 = 1$

A A A A B B B C

$y_2 = 0$

B C C C D D D D

Assume step size $\lambda = 1.0$ and $\mu = \frac{1}{4}$.

What's the update for β_C?

Example Documents (Regularized)

$$\beta[j] = (\beta[j]' + \lambda(y-p)x_i) \cdot (1-2\lambda\mu)^{m_j}$$
$$\vec{\beta} = \langle 0,0,0,0,0 \rangle$$

$y_1 = 1$

A A A A B B B C

$y_2 = 0$

B C C C D D D D

Assume step size $\lambda = 1.0$ and $\mu = \frac{1}{4}$.

What's the update for β_C?

$$\beta_C = \left(\beta_C' + \lambda \cdot (y_1 - \pi_1) \cdot x_{1,C}\right)(1 - 2 \cdot \lambda \cdot \mu)^{m_C} =$$
$$(0.0 + 1.0 \cdot (1.0 - 0.5) \cdot 1.0)\left(1 - 2 \cdot 1.0 \cdot \frac{1}{4}\right)^1$$

Example Documents (Regularized)

$$\beta[j] = (\beta[j]' + \lambda(y-p)x_i) \cdot (1 - 2\lambda\mu)^{m_j}$$
$$\vec{\beta} = \langle 0, 0, 0, 0, 0 \rangle$$

$y_1 = 1$

A A A A B B B C

$y_2 = 0$

B C C C D D D D

Assume step size $\lambda = 1.0$ and $\mu = \frac{1}{4}$.

What's the update for β_C?

$$\beta_C = \left(\beta_C' + \lambda \cdot (y_1 - \pi_1) \cdot x_{1,C}\right)(1 - 2 \cdot \lambda \cdot \mu)^{m_C} =$$
$$(0.0 + 1.0 \cdot (1.0 - 0.5) \cdot 1.0)\left(1 - 2 \cdot 1.0 \cdot \frac{1}{4}\right)^1 \quad = 0.25$$

Example Documents (Regularized)

$$\beta[j] = (\beta[j]' + \lambda(y - p)x_i) \cdot (1 - 2\lambda\mu)^{m_j}$$
$$\vec{\beta} = \langle 0, 0, 0, 0, 0 \rangle$$

$y_1 = 1$	$y_2 = 0$
A A A A B B B C	B C C C D D D D

Assume step size $\lambda = 1.0$ and $\mu = \frac{1}{4}$.

What's the update for β_D?

Example Documents (Regularized)

$$\beta[j] = (\beta[j]' + \lambda(y - p)x_i) \cdot (1 - 2\lambda\mu)^{m_j}$$
$$\vec{\beta} = \langle 0, 0, 0, 0, 0 \rangle$$

$y_1 = 1$

A A A A B B B C

$y_2 = 0$

B C C C D D D D

Assume step size $\lambda = 1.0$ and $\mu = \frac{1}{4}$.

What's the update for β_D?

We don't care: leave it for later.

Example Documents (Regularized)

$$\beta[j] = (\beta[j]' + \lambda(y-p)x_i) \cdot (1-2\lambda\mu)^{m_j}$$
$$\vec{\beta} = \langle .25, 1, 0.75, 0.25, 0 \rangle$$

$y_1 = 1$	$y_2 = 0$
A A A A B B B C	B C C C D D D D

Assume step size $\lambda = 1.0$ and $\mu = \frac{1}{4}$.

Example Documents (Regularized)

$$\beta[j] = (\beta[j]' + \lambda(y - p)x_i) \cdot (1 - 2\lambda\mu)^{m_j}$$
$$\vec{\beta} = \langle .25, 1, 0.75, 0.25, 0 \rangle$$

$y_1 = 1$	$y_2 = 0$
A A A A B B B C	B C C C D D D D

Assume step size $\lambda = 1.0$ and $\mu = \frac{1}{4}$.

Now you see the negative example. What's π_2?

Example Documents (Regularized)

$$\beta[j] = (\beta[j]' + \lambda(y-p)x_i) \cdot (1-2\lambda\mu)^{m_j}$$
$$\vec{\beta} = \langle .25, 1, 0.75, 0.25, 0 \rangle$$

$y_1 = 1$

A A A A B B B C

$y_2 = 0$

B C C C D D D D

Assume step size $\lambda = 1.0$ and $\mu = \frac{1}{4}$.

Now you see the negative example. What's π_2?

$$\pi_2 = \Pr(y_2 = 1 \,|\, \vec{x_2}) = \frac{\exp \beta^T x_i}{1+\exp \beta^T x_i} = \frac{\exp[.25+0.75+0.75+0]}{\exp[.25+0.75+0.75+0]+1} =$$

Example Documents (Regularized)

$$\beta[j] = (\beta[j]' + \lambda(y-p)x_i)\cdot(1-2\lambda\mu)^{m_j}$$
$$\vec{\beta} = \langle .25, 1, 0.75, 0.25, 0\rangle$$

$y_1 = 1$	$y_2 = 0$
A A A A B B B C	B C C C D D D D

Assume step size $\lambda = 1.0$ and $\mu = \frac{1}{4}$.

Now you see the negative example. What's π_2?

$$\pi_2 = \Pr(y_2 = 1 \,|\, \vec{x_2}) = \frac{\exp \beta^T x_i}{1+\exp \beta^T x_i} = \frac{\exp[.25+0.75+0.75+0]}{\exp[.25+0.75+0.75+0]+1} = 0.85$$

Example Documents (Regularized)

$$\beta[j] = (\beta[j]' + \lambda(y - p)x_i) \cdot (1 - 2\lambda\mu)^{m_j}$$
$$\vec{\beta} = \langle .25, 1, 0.75, 0.25, 0 \rangle$$

$y_1 = 1$	$y_2 = 0$
A A A A B B B C	B C C C D D D D

Assume step size $\lambda = 1.0$ and $\mu = \frac{1}{4}$.

$\pi_2 = 0.85$

What's the update for β_{bias}?

Example Documents (Regularized)

$$\beta[j] = (\beta[j]' + \lambda(y-p)x_i) \cdot (1-2\lambda\mu)^{m_j}$$
$$\vec{\beta} = \langle .25, 1, 0.75, 0.25, 0 \rangle$$

$y_1 = 1$	$y_2 = 0$
A A A A B B B C	B C C C D D D D

Assume step size $\lambda = 1.0$ and $\mu = \frac{1}{4}$.

What's the update for β_{bias}?

$\beta_{bias} = \left(\beta'_{bias} + \lambda \cdot (y_2 - \pi_2) \cdot x_{2,bias}\right)(1 - 2 \cdot \lambda \cdot \mu)^{m_{bias}} =$

$(0.25 + 1.0 \cdot (0.0 - 0.85) \cdot 1.0)\left(1 - 2 \cdot 1.0 \cdot \frac{1}{4}\right)^1$

Example Documents (Regularized)

$$\beta[j] = (\beta[j]' + \lambda(y-p)x_i) \cdot (1-2\lambda\mu)^{m_j}$$
$$\vec{\beta} = \langle .25, 1, 0.75, 0.25, 0 \rangle$$

$y_1 = 1$	$y_2 = 0$
A A A A B B B C	B C C C D D D D

Assume step size $\lambda = 1.0$ and $\mu = \frac{1}{4}$.

What's the update for β_{bias}?

$\beta_{bias} = \left(\beta'_{bias} + \lambda \cdot (y_2 - \pi_2) \cdot x_{2,bias}\right)(1 - 2 \cdot \lambda \cdot \mu)^{m_{bias}} =$

$(0.25 + 1.0 \cdot (0.0 - 0.85) \cdot 1.0)\left(1 - 2 \cdot 1.0 \cdot \frac{1}{4}\right)^1$ =-0.30

Example Documents (Regularized)

$$\beta[j] = (\beta[j]' + \lambda(y-p)x_i) \cdot (1-2\lambda\mu)^{m_j}$$
$$\vec{\beta} = \langle .25, 1, 0.75, 0.25, 0 \rangle$$

$y_1 = 1$

A A A A B B B C

$y_2 = 0$

B C C C D D D D

Assume step size $\lambda = 1.0$ and $\mu = \frac{1}{4}$.

What's the update for β_A?

Example Documents (Regularized)

$$\beta[j] = (\beta[j]' + \lambda(y - p)x_i) \cdot (1 - 2\lambda\mu)^{m_j}$$
$$\vec{\beta} = \langle .25, 1, 0.75, 0.25, 0 \rangle$$

$y_1 = 1$	$y_2 = 0$
A A A A B B B C	B C C C D D D D

Assume step size $\lambda = 1.0$ and $\mu = \frac{1}{4}$.

What's the update for β_A?

We don't care: leave it for later.

Example Documents (Regularized)

$$\beta[j] = (\beta[j]' + \lambda(y - p)x_i) \cdot (1 - 2\lambda\mu)^{m_j}$$
$$\vec{\beta} = \langle .25, 1, 0.75, 0.25, 0 \rangle$$

$y_1 = 1$

A A A A B B B C

$y_2 = 0$

B C C C D D D D

Assume step size $\lambda = 1.0$ and $\mu = \frac{1}{4}$.

What's the update for β_B?

Example Documents (Regularized)

$$\beta[j] = (\beta[j]' + \lambda(y - p)x_i) \cdot (1 - 2\lambda\mu)^{m_j}$$
$$\vec{\beta} = \langle .25, 1, 0.75, 0.25, 0 \rangle$$

$y_1 = 1$

A A A A B B B C

$y_2 = 0$

B C C C D D D D

Assume step size $\lambda = 1.0$ and $\mu = \frac{1}{4}$.

What's the update for β_B?

$$\beta_B = \left(\beta_B' + \lambda \cdot (y_2 - \pi_2) \cdot x_{2,B}\right)(1 - 2 \cdot \lambda \cdot \mu)^{m_B} =$$
$$(0.75 + 1.0 \cdot (0.0 - 0.85) \cdot 1.0)\left(1 - 2 \cdot 1.0 \cdot \frac{1}{4}\right)^1$$

Example Documents (Regularized)

$$\beta[j] = (\beta[j]' + \lambda(y - p)x_i) \cdot (1 - 2\lambda\mu)^{m_j}$$
$$\vec{\beta} = \langle .25, 1, 0.75, 0.25, 0 \rangle$$

$y_1 = 1$

A A A A B B B C

$y_2 = 0$

B C C C D D D D

Assume step size $\lambda = 1.0$ and $\mu = \frac{1}{4}$.

What's the update for β_B?

$\beta_B = \left(\beta_B' + \lambda \cdot (y_2 - \pi_2) \cdot x_{2,B}\right)(1 - 2 \cdot \lambda \cdot \mu)^{m_B} =$

$(0.75 + 1.0 \cdot (0.0 - 0.85) \cdot 1.0)\left(1 - 2 \cdot 1.0 \cdot \frac{1}{4}\right)^1$ =-0.05

Example Documents (Regularized)

$$\beta[j] = (\beta[j]' + \lambda(y-p)x_i) \cdot (1-2\lambda\mu)^{m_j}$$
$$\vec{\beta} = \langle .25, 1, 0.75, 0.25, 0 \rangle$$

$y_1 = 1$	$y_2 = 0$
A A A A B B B C	B C C C D D D D

Assume step size $\lambda = 1.0$ and $\mu = \frac{1}{4}$.

What's the update for β_C?

Example Documents (Regularized)

$$\beta[j] = (\beta[j]' + \lambda(y-p)x_i) \cdot (1 - 2\lambda\mu)^{m_j}$$
$$\vec{\beta} = \langle .25, 1, 0.75, 0.25, 0 \rangle$$

$y_1 = 1$

A A A A B B B C

$y_2 = 0$

B C C C D D D D

Assume step size $\lambda = 1.0$ and $\mu = \frac{1}{4}$.

What's the update for β_C?

$\beta_C = \left(\beta_C' + \lambda \cdot (y_2 - \pi_2) \cdot x_{2,C}\right)(1 - 2 \cdot \lambda \cdot \mu)^{m_C} =$

$(0.25 + 1.0 \cdot (0.0 - 0.85) \cdot 3.0)\left(1 - 2 \cdot 1.0 \cdot \frac{1}{4}\right)^1$

Example Documents (Regularized)

$$\beta[j] = (\beta[j]' + \lambda(y-p)x_i) \cdot (1-2\lambda\mu)^{m_j}$$
$$\vec{\beta} = \langle .25, 1, 0.75, 0.25, 0 \rangle$$

$y_1 = 1$	$y_2 = 0$
A A A A B B B C	B C C C D D D D

Assume step size $\lambda = 1.0$ and $\mu = \frac{1}{4}$.

What's the update for β_C?

$\beta_C = \left(\beta_C' + \lambda \cdot (y_2 - \pi_2) \cdot x_{2,C}\right)(1 - 2 \cdot \lambda \cdot \mu)^{m_C} =$

$(0.25 + 1.0 \cdot (0.0 - 0.85) \cdot 3.0)\left(1 - 2 \cdot 1.0 \cdot \frac{1}{4}\right)^1$ =-1.15

Example Documents (Regularized)

$$\beta[j] = (\beta[j]' + \lambda(y - p)x_i) \cdot (1 - 2\lambda\mu)^{m_j}$$
$$\vec{\beta} = \langle .25, 1, 0.75, 0.25, 0 \rangle$$

$y_1 = 1$	$y_2 = 0$
A A A A B B B C	B C C C D D D D

Assume step size $\lambda = 1.0$ and $\mu = \frac{1}{4}$.

What's the update for β_D?

Example Documents (Regularized)

$$\beta[j] = (\beta[j]' + \lambda(y-p)x_i) \cdot (1 - 2\lambda\mu)^{m_j}$$
$$\vec{\beta} = \langle .25, 1, 0.75, 0.25, 0 \rangle$$

$y_1 = 1$	$y_2 = 0$
A A A A B B B C	B C C C D D D D

Assume step size $\lambda = 1.0$ and $\mu = \frac{1}{4}$.

What's the update for β_D?

$\beta_D = \left(\beta'_D + \lambda \cdot (y_2 - \pi_2) \cdot x_{2,D}\right)(1 - 2 \cdot \lambda \cdot \mu)^{m_D} =$

$(0.0 + 1.0 \cdot (0.0 - 0.85) \cdot 4.0)\left(1 - 2 \cdot 1.0 \cdot \frac{1}{4}\right)^2$

Example Documents (Regularized)

$$\beta[j] = (\beta[j]' + \lambda(y - p)x_i) \cdot (1 - 2\lambda\mu)^{m_j}$$
$$\vec{\beta} = \langle .25, 1, 0.75, 0.25, 0 \rangle$$

$y_1 = 1$	$y_2 = 0$
A A A A B B B C	B C C C D D D D

Assume step size $\lambda = 1.0$ and $\mu = \frac{1}{4}$.

What's the update for β_D?

$\beta_D = \left(\beta_D' + \lambda \cdot (y_2 - \pi_2) \cdot x_{2,D}\right)(1 - 2 \cdot \lambda \cdot \mu)^{m_D} =$

$(0.0 + 1.0 \cdot (0.0 - 0.85) \cdot 4.0)\left(1 - 2 \cdot 1.0 \cdot \frac{1}{4}\right)^2$ =-0.85

Next time . . .

- Multinomial logistic regression in sklearn (more than one option)
- Crafting effective features
- Preparation for third homework