Computational Linguistics

Natural Language Processing

University of Maryland

Classification Examples
Reminder: Logistic Regression

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

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

- Discriminative prediction: \( p(y|x) \)
- Classification uses: ad placement, spam detection
- What we didn’t talk about is how to learn \( \beta \) from data
Logistic Regression: Objective Function

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

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

1. Initialize a vector $B$ to be all zeros
2. For $t = 1, \ldots, T$
   ▶ For each example $\tilde{x}_i, y_i$ and feature $j$:
     ▶ Compute $\pi_i \equiv \Pr(y_i = 1 | \tilde{x}_i)$
     ▶ Set $\beta[j] = \beta[j]' + \lambda (y_i - \pi_i)x_{i,j}$
3. Output the parameters $\beta_1, \ldots, \beta_d$. 
You first see the positive example. First, compute $\pi_1$
\[ \beta[j] = \beta[j] + \lambda (y_i - \pi_i)x_i \]
\[ \vec{\beta} = \langle 0, 0, 0, 0, 0 \rangle \]

You first see the positive example. First, compute \( \pi_1 \)

\[ \pi_1 = \Pr(y_1 = 1 | \vec{x}_1) = \frac{\exp \beta^T x_i}{1 + \exp \beta^T x_i} = \]

\( y_1 = 1 \)

A A A A B B B C

\( y_2 = 0 \)

B C C C D D D D

(Assume step size \( \lambda = 1.0 \).)
Example Documents

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

\[ \vec{\beta} = \langle 0, 0, 0, 0, 0 \rangle \]

\( y_1 = 1 \)

A A A A B B B C

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

\( y_2 = 0 \)

B C C C D D D D D

You first see the positive example. First, compute \( \pi_1 \)

\[ \pi_1 = \Pr(y_1 = 1 | \vec{x}_1) = \frac{\exp \beta^T x_i}{1 + \exp \beta^T x_i} = \frac{\exp 0}{\exp 0 + 1} = 0.5 \]
\[ \beta[j] = \beta[j] + \lambda(y_i - \pi_i) x_i \]
\[ \bar{\beta} = \langle 0, 0, 0, 0, 0 \rangle \]

\[ y_1 = 1 \quad y_2 = 0 \]

A A A A B B B C
B C C C D D D D

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

\( \pi_1 = 0.5 \) What's the update for \( \beta_{bias} \)?
\[
\beta[j] = \beta[j] + \lambda(y_i - \pi_i)x_i \\
\hat{\beta} = \langle 0, 0, 0, 0, 0 \rangle
\]

\(y_1 = 1\)

A A A A B B B C

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

\(y_2 = 0\)

B C C C D D D D D D

What’s the update for \(\beta_{bias}\)?

\[
\beta_{bias} = \beta'_{bias} + \lambda \cdot (y_1 - \pi_1) \cdot x_{1,bias} = 0.0 + 1.0 \cdot (1.0 - 0.5) \cdot 1.0
\]
\[
\beta[j] = \beta[j] + \lambda(y_i - \pi_i)x_i
\]
\[
\vec{\beta} = \langle 0, 0, 0, 0, 0 \rangle
\]

\(y_1 = 1\)

A A A A B B B C

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

\(y_2 = 0\)

B C C C D D D D D D

What’s the update for \(\beta_{bias}\)?

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

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

\( y_1 = 1 \)
A A A A B B B C

\( y_2 = 0 \)
B C C C D D D D

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

What’s the update for \( \beta_A \)?
\[ \beta[j] = \beta[j] + \lambda(y_i - \pi_i)x_i \]

\[ \vec{\beta} = \langle 0, 0, 0, 0, 0 \rangle \]

\begin{align*}
\text{y}_1 &= 1 \\
\text{A A A A B B B C} & \quad \text{(Assume step size } \lambda = 1.0.\text{)}
\end{align*}

\begin{align*}
\text{y}_2 &= 0 \\
\text{B C C C D D D D} & \quad \text{D D D D}
\end{align*}

What’s the update for \( \beta_A \)?

\[ \beta_A = \beta'_A + \lambda \cdot (y_1 - \pi_1) \cdot x_{1,A} = 0.0 + 1.0 \cdot (1.0 - 0.5) \cdot 4.0 \]
\[ \beta[j] = \beta[j] + \lambda(y_i - \pi_i)x_i \]

\[ \hat{\beta} = \langle 0, 0, 0, 0, 0 \rangle \]

\[ y_1 = 1 \]

A A A A B B B C

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

\[ y_2 = 0 \]

B C C C D D D D D

What’s the update for \( \beta_A \)?

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

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

\( y_1 = 1 \)

A A A A B B B C

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

\( y_2 = 0 \)

B C C C D D D D D

What’s the update for \( \beta_B \)?
\[ \beta[j] = \beta[j] + \lambda(y_i - \pi_i)x_i \]
\[ \vec{\beta} = (0, 0, 0, 0, 0) \]

\( y_1 = 1 \)  
A A A A B B B C  
(Assume step size \( \lambda = 1.0 \).)

\( y_2 = 0 \)  
B C C C D D D D D D

What’s the update for \( \beta_B \)?
\[ \beta_B = \beta'_B + \lambda \cdot (y_1 - \pi_1) \cdot x_{1,B} = 0.0 + 1.0 \cdot (1.0 - 0.5) \cdot 3.0 \]
$\beta[j] = \beta[j] + \lambda(y_i - \pi_i)x_i$

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

$y_1 = 1$

A A A A B B B C

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

$y_2 = 0$

B C C C D D D D D

What's the update for $\beta_B$?

$\beta_B = \beta'_B + \lambda \cdot (y_1 - \pi_1) \cdot x_{1,B} = 0.0 + 1.0 \cdot (1.0 - 0.5) \cdot 3.0 = 1.5$
\[ \beta[j] = \beta[j] + \lambda(y_i - \pi_i)x_i \]
\[ \vec{\beta} = \langle 0, 0, 0, 0, 0 \rangle \]

\[ y_1 = 1 \]
A A A A B B B C
(Assume step size \( \lambda = 1.0 \).)

\[ y_2 = 0 \]
B C C C D D D D D

What’s the update for \( \beta_C \)?
\[ \beta[j] = \beta[j] + \lambda(y_i - \pi_i)x_i \]
\[ \vec{\beta} = \langle 0, 0, 0, 0, 0 \rangle \]

\( y_1 = 1 \)
A A A A B B B C

(Imagine step size \( \lambda = 1.0 \).)

\( y_2 = 0 \)
B C C C D D D D D D

What’s the update for \( \beta_C \)?
\[ \beta_C = \beta'_C + \lambda \cdot (y_1 - \pi_1) \cdot x_{1,C} = 0.0 + 1.0 \cdot (1.0 - 0.5) \cdot 1.0 \]
Example Documents

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

\[ \hat{\beta} = \langle 0, 0, 0, 0, 0 \rangle \]

\( y_1 = 1 \)

A A A A B B B C

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

\( y_2 = 0 \)

B C C C D D D D

What’s the update for \( \beta_C \)?

\[ \beta_C = \beta_C' + \lambda \cdot (y_1 - \pi_1) \cdot x_{1,C} = 0.0 + 1.0 \cdot (1.0 - 0.5) \cdot 1.0 = 0.5 \]
\[ \beta[j] = \beta[j] + \lambda(y_i - \pi_i)x_i \]
\[ \vec{\beta} = \langle 0, 0, 0, 0, 0 \rangle \]

\( y_1 = 1 \)

A A A A B B B C

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

\( y_2 = 0 \)

B C C C D D D D D D

What’s the update for \( \beta_D \)?
Example Documents

\[ \beta[j] = \beta[j] + \lambda(y_i - \pi_i)x_i \]
\[ \hat{\beta} = (0, 0, 0, 0, 0) \]

\[ y_1 = 1 \]
\[ \text{A A A A B B B C} \]
(Assume step size \( \lambda = 1.0 \).)

\[ y_2 = 0 \]
\[ \text{B C C C D D D D D D} \]

What’s the update for \( \beta_D \)?
\[ \beta_D = \beta'_D + \lambda \cdot (y_1 - \pi_1) \cdot x_{1,D} = 0.0 + 1.0 \cdot (1.0 - 0.5) \cdot 0.0 \]
\[ \beta[j] = \beta[j] + \lambda(y_i - \pi_i)x_i \]
\[ \bar{\beta} = (0, 0, 0, 0, 0) \]

\( y_1 = 1 \)
\begin{align*}
A & \quad A & \quad A & \quad A & \quad B & \quad B & \quad B & \quad C \\
\end{align*}

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

\( y_2 = 0 \)
\begin{align*}
B & \quad C & \quad C & \quad C & \quad D & \quad D & \quad D & \quad D & \quad D
\end{align*}

What’s the update for \( \beta_D \)?
\[ \beta_D = \beta'_D + \lambda \cdot (y_1 - \pi_1) \cdot x_{1,D} = 0.0 + 1.0 \cdot (1.0 - 0.5) \cdot 0.0 = 0.0 \]
Example Documents

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

\( y_1 = 1 \)
A A A A B B B C
(Assume step size \( \lambda = 1.0 \).)

\( y_2 = 0 \)
B C C C D D D D D D

Now you see the negative example. What's \( \pi_2 \)?
$$\beta[j] = \beta[j] + \lambda(y_i - \pi_i)x_i$$
$$\vec{\beta} = \langle .5, 2, 1.5, 0.5, 0 \rangle$$

$$y_1 = 1$$
A A A A B B B C
(Assume step size $\lambda = 1.0$.)

$$y_2 = 0$$
B C C C D D D D D D

Now you see the negative example. What's $\pi_2$?

$$\pi_2 = \Pr(y_2 = 1 | \vec{x}_2) = \frac{\exp \beta^T x_i}{1 + \exp \beta^T x_i} = \frac{\exp{.5 + 1.5 + 1.5 + 0}}{\exp{.5 + 1.5 + 1.5 + 0} + 1} =$$
$\beta[j] = \beta[j] + \lambda(y_i - \pi_i)x_i$

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

$y_1 = 1$

A A A A B B B C

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

$y_2 = 0$

B C C C D D D D D

Now you see the negative example. What’s $\pi_2$?

$\pi_2 = \Pr(y_2 = 1 | \mathbf{x}_2) = \frac{\exp \mathbf{\beta}^T \mathbf{x}_i}{1 + \exp \mathbf{\beta}^T \mathbf{x}_i} = \frac{\exp \{.5 + 1.5 + 1.5 + 0\}}{\exp \{.5 + 1.5 + 1.5 + 0\} + 1} = 0.97$
\[ \beta[j] = \beta[j] + \lambda(y_i - \pi_i)x_i \]
\[ \vec{\beta} = (0.5, 2, 1.5, 0.5, 0) \]

\( y_1 = 1 \)
A A A A B B B C

\( y_2 = 0 \)
B C C C D D D D D D D D

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

Now you see the negative example. What's \( \pi_2 \)?
\( \pi_2 = 0.97 \)
What's the update for \( \beta_{bias} \)?
\[ \beta[j] = \beta[j] + \lambda(y_i - \pi_i)x_i \]
\[ \tilde{\beta} = \langle .5, 2, 1.5, 0.5, 0 \rangle \]

\[
y_1 = 1
\]
A A A A B B B C

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

\[
y_2 = 0
\]
B C C C D D D D D D D

What’s the update for \( \beta_{bias} \)?
\[
\beta_{bias} = \beta'_{bias} + \lambda \cdot (y_2 - \pi_2) \cdot x_2, bias = 0.5 + 1.0 \cdot (0.0 - 0.97) \cdot 1.0
\]
\[ \beta[j] = \beta[j] + \lambda(y_i - \pi_i)x_i \]
\[ \vec{\beta} = \langle .5, 2, 1.5, 0.5, 0 \rangle \]

\[ y_1 = 1 \]
A A A A B B B C

\[ y_2 = 0 \]
B C C C D D D D D

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

What's the update for \( \beta_{bias} \)?
\[ \beta_{bias} = \beta'_{bias} + \lambda \cdot (y_2 - \pi_2) \cdot x_{2,bias} = 0.5 + 1.0 \cdot (0.0 - 0.97) \cdot 1.0 = -0.47 \]
Example Documents

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

\[ \vec{\beta} = (.5, 2, 1.5, 0.5, 0) \]

\[ y_1 = 1 \]

A A A A B B B C

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

\[ y_2 = 0 \]

B C C C D D D D D D D

What’s the update for \( \beta_A \)?
\[ \beta[j] = \beta[j] + \lambda(y_i - \pi_i)x_i \]
\[ \hat{\beta} = (.5, 2, 1.5, 0.5, 0) \]

\[ y_1 = 1 \]
A A A A B B B C
(Assume step size \( \lambda = 1.0 \).)

\[ y_2 = 0 \]
B C C C D D D D D D

What’s the update for \( \beta_A \)?
\[ \beta_A = \beta_A' + \lambda \cdot (y_2 - \pi_2) \cdot x_{2,A} = 2.0 + 1.0 \cdot (0.0 - 0.97) \cdot 0.0 \]
$\beta[j] = \beta[j] + \lambda (y_i - \pi_i) x_i$

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

$y_1 = 1$

A A A A B B B C

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

$y_2 = 0$

B C C C D D D D D D

What’s the update for $\beta_A$?

$\beta_A = \beta'_A + \lambda \cdot (y_2 - \pi_2) \cdot x_{2,A} = 2.0 + 1.0 \cdot (0.0 - 0.97) \cdot 0.0 = 2.0$
β[j] = β[j] + λ(y_i − π_i)x_i

β = ⟨.5, 2, 1.5, 0.5, 0⟩

y_1 = 1
A A A A B B B C
(Assume step size \( \lambda = 1.0 \).)

y_2 = 0
B C C C D D D D D D

What’s the update for β_B?
\[ \beta[j] = \beta[j] + \lambda (y_i - \pi_i) x_i \]

\[ \hat{\beta} = \langle .5, 2, 1.5, 0.5, 0 \rangle \]

\[ y_1 = 1 \]
\[ y_2 = 0 \]

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

What’s the update for \( \beta_B \)?

\[ \beta_B = \beta'_B + \lambda \cdot (y_2 - \pi_2) \cdot x_{2,B} = 1.5 + 1.0 \cdot (0.0 - 0.97) \cdot 1.0 \]
\[ \beta[j] = \beta[j] + \lambda(y_i - \pi_i)x_i \]
\[ \hat{\beta} = \langle .5, 2, 1.5, 0.5, 0 \rangle \]

\[ y_1 = 1 \]
A A A A B B B C
(Assume step size \( \lambda = 1.0 \).)

\[ y_2 = 0 \]
B C C C D D D D

What’s the update for \( \beta_B \)?
\[ \beta_B = \beta_B' + \lambda \cdot (y_2 - \pi_2) \cdot x_{2,B} = 1.5 + 1.0 \cdot (0.0 - 0.97) \cdot 1.0 = 0.53 \]
\[ \beta[j] = \beta[j] + \lambda(y_i - \pi_i)x_i \]

\[ \vec{\beta} = (0.5, 2, 1.5, 0.5, 0) \]

<table>
<thead>
<tr>
<th>(y_1 = 1)</th>
<th>(y_2 = 0)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A A A A B B B C</td>
<td>B C C C D D D D</td>
</tr>
</tbody>
</table>

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

What’s the update for \(\beta_C\)?
\[ \beta[j] = \beta[j] + \lambda(y_i - \pi_i)x_i \]
\[ \tilde{\beta} = \langle .5, 2, 1.5, 0.5, 0 \rangle \]

\[ y_1 = 1 \]
A A A A B B B C

\[ y_2 = 0 \]
B C C C D D D D D D

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

What’s the update for \( \beta_C \)?
\[ \beta_C = \beta'_C + \lambda \cdot (y_2 - \pi_2) \cdot x_{2,C} = 0.5 + 1.0 \cdot (0.0 - 0.97) \cdot 3.0 \]
\[ \beta[j] = \beta[j] + \lambda(y_i - \pi_i)x_i \]
\[ \hat{\beta} = (0.5, 2, 1.5, 0.5, 0) \]

\( y_1 = 1 \)
\[
\text{A A A A B B B C}
\]

(assume step size \( \lambda = 1.0 \).)

\( y_2 = 0 \)
\[
\text{B C C C D D D D D D}
\]

what's the update for \( \beta_C \)?

\[ \beta_C = \beta_C' + \lambda \cdot (y_2 - \pi_2) \cdot x_{2,C} = 0.5 + 1.0 \cdot (0.0 - 0.97) \cdot 3.0 = -2.41 \]
Example Documents

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

\[ \vec{\beta} = \langle 0.5, 2, 1.5, 0.5, 0 \rangle \]

\[ y_1 = 1 \]

A A A A B B B C

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

\[ y_2 = 0 \]

B C C C D D D D D D D D

What’s the update for \( \beta_D \)?
\[ \beta[j] = \beta[j] + \lambda (y_i - \pi_i) x_i \]
\[ \hat{\beta} = \langle .5, 2, 1.5, 0.5, 0 \rangle \]

\[ y_1 = 1 \]
A A A A B B B C

\[ y_2 = 0 \]
B C C C D D D D D D

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

What’s the update for \( \beta_D \)?
\[ \beta_D = \beta_D' + \lambda \cdot (y_2 - \pi_2) \cdot x_{2,D} = 0.0 + 1.0 \cdot (0.0 - 0.97) \cdot 4.0 \]
Example Documents

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

\[ y_1 = 1 \]
A A A A B B B C
(Assume step size \( \lambda = 1.0 \).)

\[ y_2 = 0 \]
B C C C D D D D

What’s the update for \( \beta_D \)?
\[ \beta_D = \beta'_D + \lambda \cdot (y_2 - \pi_2) \cdot x_{2,D} = 0.0 + 1.0 \cdot (0.0 - 0.97) \cdot 4.0 = -3.88 \]
Different Activation Function

Your boss demands that you replace the sigmoid function in logistic regression with the trigonometric sin function because it looks the same and he has a sin button on his calculator.

1. Plot both between -1 and 1. What would a choice of constants $A$ and $B$ be that would make $s(z) = A \sin(Bz) + C$ look as much like the logistic function?
2. What would be the update for an example? How is it different?
3. Would there be any other problems with using this formulation?
What does “look like” mean?

- Same min and max
- Derivative at 0 should match
- Same value at 0
What does “look like” mean?

- Same min and max
- Derivative at 0 should match
- Same value at 0

\[ s(z) \equiv \frac{1}{2} \sin \left( \frac{z}{2} \right) + \frac{1}{2} \]  \hspace{1cm} (5)
What does “look like” mean?

- Same min and max: \( s(\pi) = \frac{1}{2} \sin\left(\frac{\pi}{2}\right) + \frac{1}{2} = \frac{1}{2} + \frac{1}{2} = 1 \)
- Derivative at 0 should match
- Same value at 0

\[
s(z) \equiv \frac{1}{2} \sin\left(\frac{z}{2}\right) + \frac{1}{2}
\]  
(5)
What does “look like” mean?

- Same min and max: 
  \[ s(\pi) = \frac{1}{2} \sin\left(\frac{\pi}{2}\right) + \frac{1}{2} = \frac{1}{2} + \frac{1}{2} = 1 \]

- Derivative at 0 should match

- Same value at 0

\[ s(z) \equiv \frac{1}{2} \sin\left(\frac{z}{2}\right) + \frac{1}{2} \quad (5) \]
What does “look like” mean?

• Same \( \min \) and \( \max \): \( s(\pi) = \frac{1}{2} \sin \left( \frac{\pi}{2} \right) + \frac{1}{2} = \frac{1}{2} + \frac{1}{2} = 1 \)

• Derivative at 0 should match

• Same value at 0: \( s(0) = \frac{1}{2} \sin 0 + \frac{1}{2} = 0 + \frac{1}{2} = \frac{1}{2} \)

\[
s(z) \equiv \frac{1}{2} \sin \left( \frac{z}{2} \right) + \frac{1}{2} \quad (5)
\]
What does “look like” mean?

- **Same min and max:** 
  \[ s(\pi) = \frac{1}{2} \sin\left(\frac{\pi}{2}\right) + \frac{1}{2} = \frac{1}{2} + \frac{1}{2} = 1 \]

- **Derivative at 0 should match:** 
  \[ s'(0) = \frac{1}{4} \cos 0 = \frac{1}{4} \]

- **Same value at 0:** 
  \[ s(0) = \frac{1}{2} \sin 0 + \frac{1}{2} = 0 + \frac{1}{2} = \frac{1}{2} \]

\[
s(z) \equiv \frac{1}{2} \sin\left(\frac{z}{2}\right) + \frac{1}{2} \quad (5)
\]
Perfect match!
What can go wrong?

Can you think of two different examples that have the same update but shouldn’t?
What can go wrong?

Can you think of two different examples that have the same update but shouldn’t?
The update becomes a function of sine and cosine. If the input the function is near zero, everything is fine, but beyond that, as the gradient is a periodic function (every $\frac{2\pi}{\beta}$ you’ll repeat), the gradient is possible to be zero or inf.
If you are interested ...

Let us define:

\[ \pi_i = \frac{1}{2} \sin \left( \frac{\beta^T x_i}{2} \right) + \frac{1}{2} \]

Thus,

\[ \frac{\partial \pi_i}{\partial \beta_j} = \frac{1}{4} \cos \left( \frac{\beta^T x_i}{2} \right) x_{i,j} \tag{6} \]
To ease notation, let us further define:

\[ z' = \frac{\mathbf{\beta}^T \mathbf{x}_i}{2} \]

Thus,

\[
\frac{\partial \mathcal{L}_i}{\partial \beta_j} = \begin{cases} 
\frac{1}{\pi_i} \frac{\partial \pi_i}{\partial \beta_j} & \text{if } y_i = 1 \\
\frac{1}{1-\pi_i} \left( -\frac{\partial \pi_i}{\partial \beta_j} \right) & \text{if } y_i = 0
\end{cases} = \begin{cases} 
\frac{\cos z' }{2(\sin z' + 1)} x_{i,j} & \text{if } y_i = 1 \\
\frac{\cos z' }{2(\sin z' - 1)} x_{i,j} & \text{if } y_i = 0
\end{cases}
\] (7)