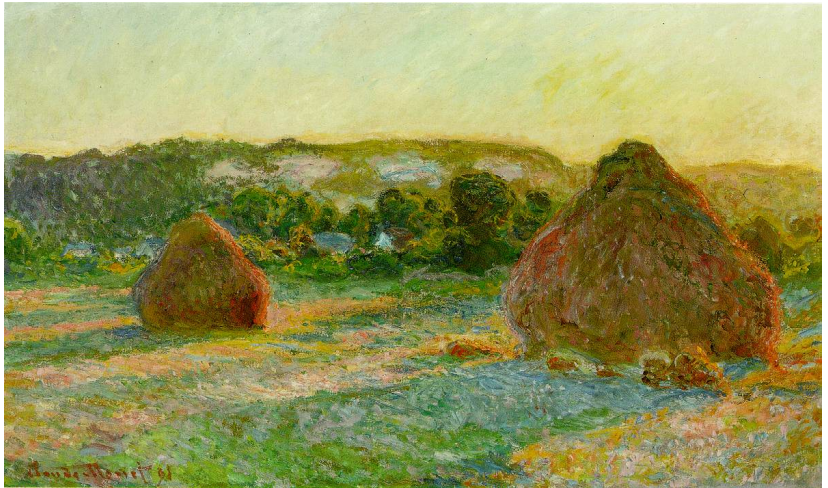


Topic Models

Guest: Xanda Schofield

University of Maryland

Introduction



Non-literal Poetry (Lisa Rhody)



Topic 2	Poem Title
0.535248643	When to the sessions of sweet silent thought (Sonnet 30)
0.533343438	By ways remote and distant waters sped (101)
0.517398877	A Psalm of Life
0.481152152	We Wear the Mask
0.477938906	The times are nightfall, look, their light grows less
0.472091675	The Slave's Complaint
0.451175606	The Guitar
0.447100571	Tears in Sleep
0.446314271	The Man with the Hoe
0.437962153	A Short Testament
0.433767746	Beyond the Years
0.433152279	Dead Fires
0.429638773	O Little Root of a Dream
0.427326132	Bangladesh II
0.425835136	Vitae Summa Brevis Spem Nos Vetat Incohare Longam

Health Information Needs (Rediet Abebe)

Topic	20 Most Representative Words	Sample Queries from Top 100
<i>Symptoms</i> (2.28%)	pain, sign, lymph, swollen, nodes, sore, symptom, symptoms, throat, infection, body, back, positive, pains, stomach, fever, neck, headache, glands, patient	hiv painfull jaw hiv swollen lymph nodes hiv swollen gland throat joints pain hiv
<i>Natural Cure</i> (0.74%)	cure, oil, black, healing, heal, healed, seed, herbs, natural, cures, moringa, kill, cured, testimonials, coconut, traditional, god, garlic, lemon, aloe	prophet bushiri hiv miracles hiv garlic lemon honey coloidal silver hiv testimonials
<i>Epidemiology</i> (0.59%)	statistics, report, 2015, global, unaids, 2016, united, epidemic, besigye, kizza, children, 2014, progress, 2010, response, nations, nigeria, prevalence, million, sa	unaids global aids report mia khalifa hiv hiv 2030
<i>Drugs</i> (0.85%)	drug, treatment, patients, abuse, therapy, drugs, resistance, antiretroviral, substance, adherence, alcohol, art, failure, spread, leads, patient, relationship, transmission, effect	stanford hiv drug resistance hiv drug therapy resistance virological failure hiv



Non-text Applications of Topic Models

- Genes
- Music
- Images
- Items (e.g., in archeology)
- Movies (for recommendations)

THIS IS YOUR MACHINE LEARNING SYSTEM?

YUP! YOU POUR THE DATA INTO THIS BIG PILE OF LINEAR ALGEBRA, THEN COLLECT THE ANSWERS ON THE OTHER SIDE.

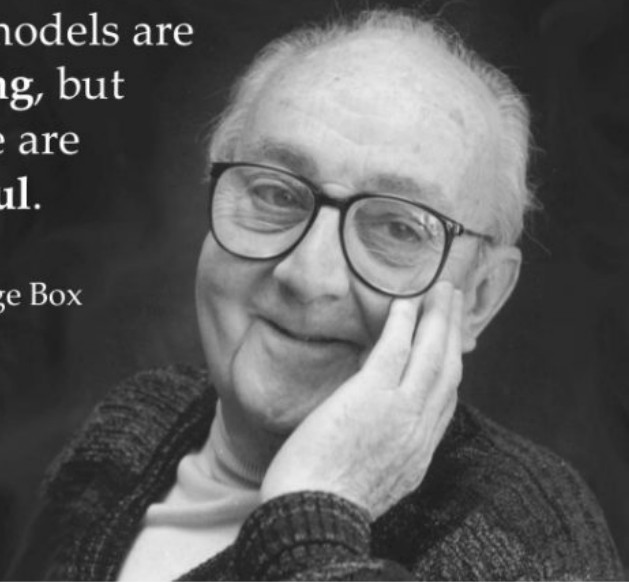
WHAT IF THE ANSWERS ARE WRONG?

JUST STIR THE PILE UNTIL THEY START LOOKING RIGHT.



All models are
wrong, but
some are
useful.

- George Box



Inference: Going from Data to Model

Gibbs Sampling

- Simpler
- Can be faster

Variational Inference

- More broadly applicable
- Can parallelize
- Fits with Pytorch



Suggested toolkits:

- Mallet (Java, Gibbs)
- Gensim (Python, variational)