Midterm: wed, March 4

open book, notes

absolutely no communication – wireless or other

Machine Learning 10-601

Tom M. Mitchell
Machine Learning Department
Carnegie Mellon University

March 2, 2009

Midterm Course Correction Survey* (5 min)

1. if you could change just one thing to improve the course, what should we do?

2. if you could pick just one thing to keep because it's helpful, what is that?

* anonymous survey

Machine Learning:

Study of algorithms that

- improve their performance P
- at some task T → learn $f: X \Rightarrow Y$ or $P(Y|X, \theta)$
- with experience E → precollect data with fully obs vars. (EM). incompletely obs)

best fitting data (MLE for θ)

best fitting (data + priors) MAP

Some topics we've covered:

- Decision trees
  - overfitting, pruning
  - information gain, entropy
- Basic probability
  - MLE's, MAP
  - manipulating probabilities
  - Bayes rule
- Naïve Bayes
  - conditional indep.
  - Discrete, Gaussian
- Logistic regression
  - derive from Naïve Bayes
  - Conditional likelihood
  - gradient descent
- Logistic regression
  - derive from Naïve Bayes
  - Conditional likelihood
  - MLE, MAP, regularization
  - gradient descent
- Cross validation, feature selection, regularization,
- Linear regression
  - probabilistic interpretation
- Bayes nets
  - representation of joint distrib.
  - cond. indep./ D-separation
  - simple inference
  - learning from fully observed data
  - learning using EM
  - clustering using EM

learn $f: X \to Y$ or $P(Y|X)$

Given a learning task.

1. pick a representation (parameterized θ) — model class / hypothesis space

2. pick an objective fn.

det. $\epsilon \sim N(0,\sigma)$ fit the data — ML likelihood

$Y = f(x) + \epsilon$

$$\theta \leftarrow \arg\max_\theta \prod_{l \text{ examples}} P(x^l y^l | \theta)$$

MAP

$$\theta \leftarrow \arg\max_\theta P(\theta | \{x^l, y^l\}) \propto \prod_l P(x^l y^l | \theta) \cdot P(\theta)$$

$$MCLE = \arg\max_\theta \prod_l P(Y^l | X^l \theta)$$

Max Likelihood

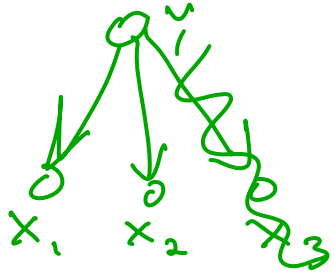
obs W

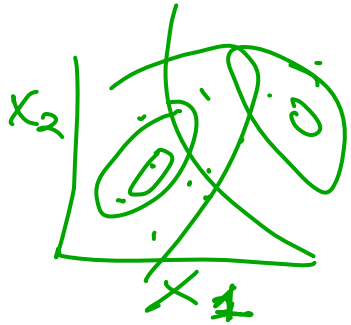
$\neg$obs Z

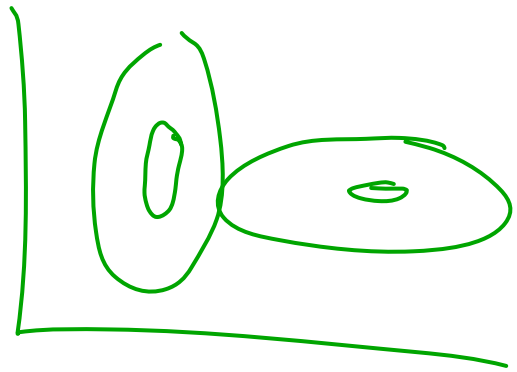
$$\arg\max_\theta \sum_{P(z|w)} \left[\prod_l P(w^l, z^l | \theta) \right] \leftarrow EM$$

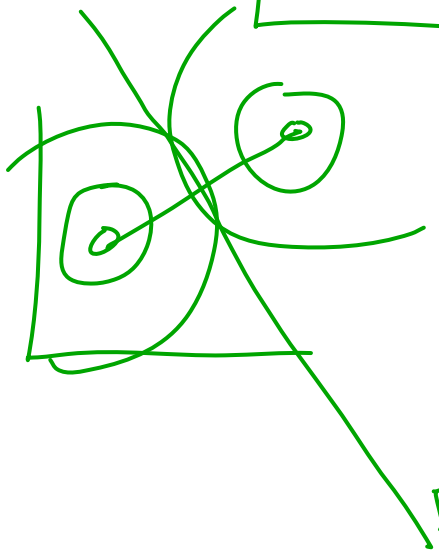
Continuous

GNB

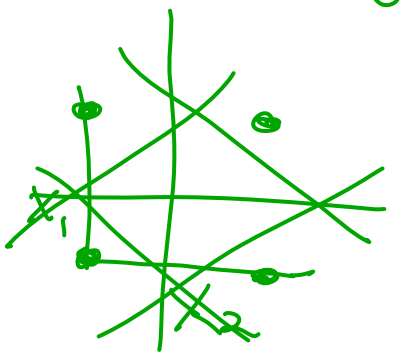


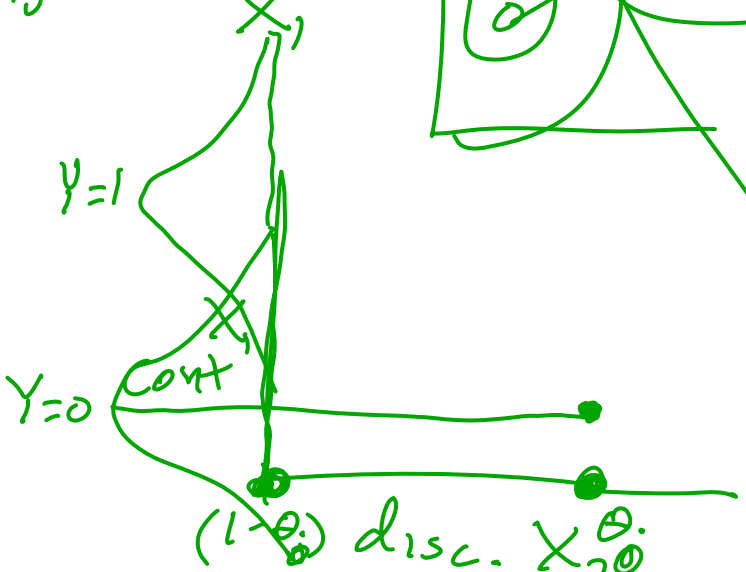






if $\sigma_{ij} = \sigma$





$P(X_i | Y_j)$

$P(Y|X) \propto P(Y) \prod_i P(X_i|Y)$

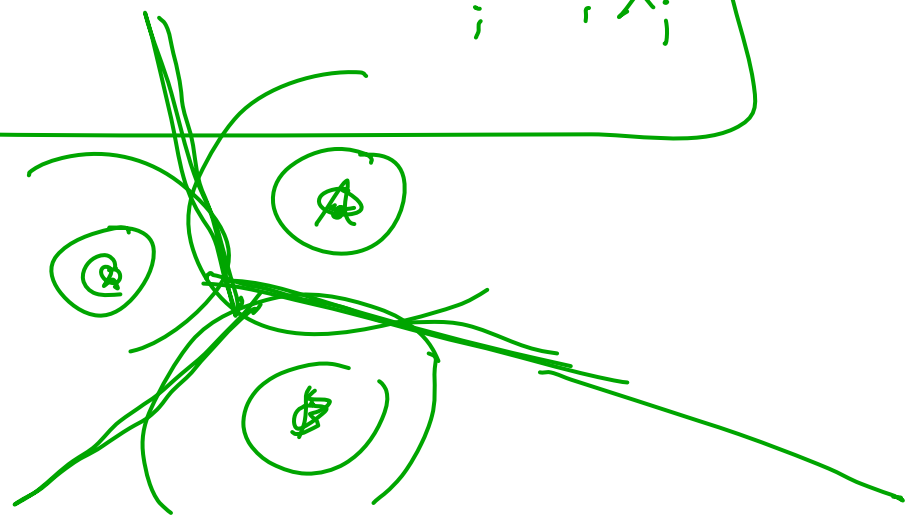
$$\ln P(Y|X) = \text{---} + \sum_i \ln\left(\frac{P(x_i|Y=1)}{P(x_i|Y=0)}\right)$$

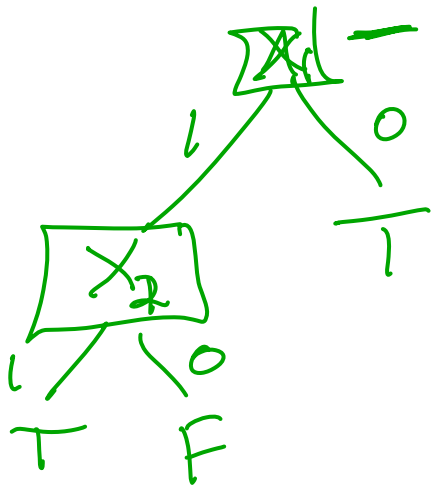
log reg $P(Y=1|X) = \frac{1}{1+\exp(\sum_i w_i x_i)}$

$P(Y=1|X) = \frac{\exp(\cdot)}{1+\exp(\cdot)}$

$$\ln \frac{P(Y=1|X)}{P(Y=0|X)} = \sum_i w_i x_i$$

$$\ln\left(\frac{P(Y=1|X)}{P(Y\neq 1|X)}\right)$$




X_1
1
0
X_2
1
0
T
F
T


1
X_2
0
0
X_1
1