Estimating Probabilities:

Slide 1:

Last week we saw:

- We can estimate probabilities first to build a model (without data first)
- Sometimes estimating these probabilities is harder than what we want to solve (like in a classification problem)
- See Slides for more benefits/drawbacks of estimating probabilities first

Slide 2-5:

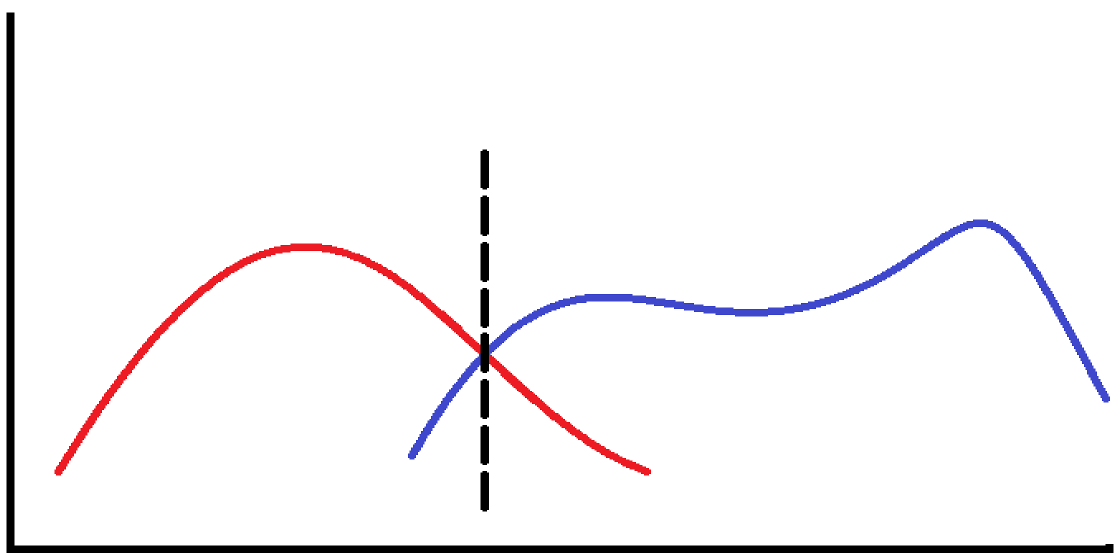
- See Slides

Slide 6:

- Note, this derivation/result (cumulative distribution) is different than slide 5 because we do not want a single probability

Slide 7-8:

- Estimating density is almost impossible, This is a problem when trying to classify
- Example, we need to compare density to find the point in the image below



Slide 9-10:

- See Slides

Slide 11:

- Note the trick in this slide. It is easier in this case to work with $\gamma = 1 / \sigma$

Slide 12:

- See Slides

–

–

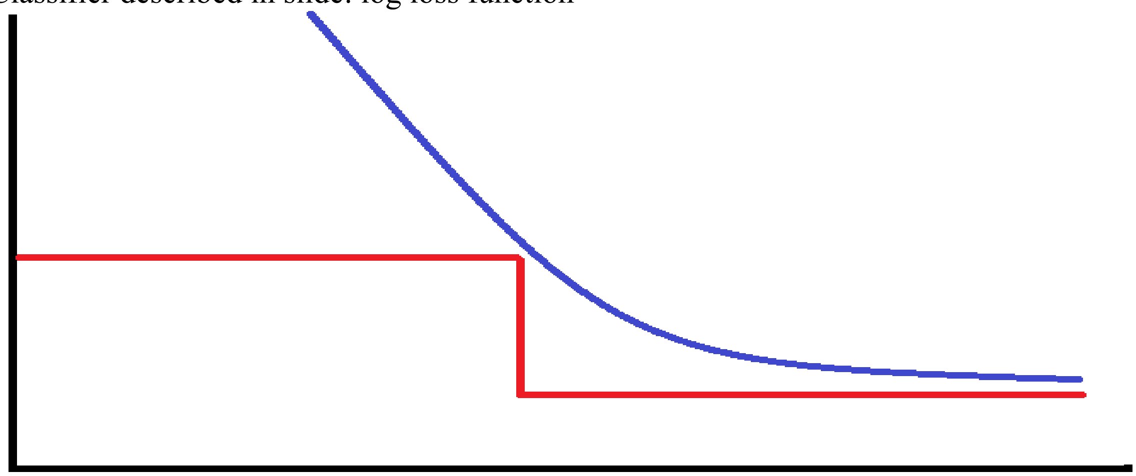
–

Slide 13:

- Remember, there are two ways to write P(x,y)
  - See Slides for
- Discriminative approach is useful in classifiers who don't care about P(X), so we can allow this to drop out of the equation
- Note there is a difference in the Discriminative and Generative normalization equations

Slide 14:

- Desired classification: max(0, 1-z) – RED
- Classifier described in slide: log loss function



Slide 15-17:

- See Slides

Slide 18:

- Note to get an unbiased estimator for variance, divide by (n-1) instead of (n)

Slide 19:

- Unbiased estimators are not a good idea if we have a priori information

Slide 20-21:

- See slides

Slide 22:

- Assume $0 \leq \Theta \leq 1$

Slide 23:

- Good initial values for α, β are ½ or 1/3

Slide 24:

- Averaging is a good idea for small data sets, but computation gets difficult for large data sets

–

Slide 25:

- Both views have capacity control (initial bias)
- Summary
  - Estimating probabilities is often a good idea
  - Estimating densities is near impossible
    - no one is really sure with continuous probabilities