Few words about midterm

Stats

- Max: 98
- Mean: 70.89
- Median: 74.5
- Std: 17.9

If you have grading questions, please bring it to me by the end of this week (Friday 5pm)

Main Problems

- Modeling assumptions of LDA and Logistic regression
  - LDA: gaussian class conditional distribution, shared covariance matrix -> $p(y|X) \sim$ logistic function
  - But the other way is not true – LDA makes stronger assumptions
- Naïve Bayes
  - Wrong direction $P(A|y)$, not $P(y|A)$
  - Conditional independence assumption

SVM: structural risk minimization

- Structural risk minimization:

$$\varepsilon \leq \varepsilon_{\mathrm{T}} + \sqrt{\frac{d(\log\frac{2m}{d}+1)+\log\frac{4}{\delta}}{m}}$$

- Softmargin SVM:
  - Slack variables can be viewed as errors (first term)
  - Maximizing Margin -> reducing the vc-dimension (second term)
  - Trading off between these two -> structural risk minimization

- Unbounded Decision tree: vc-dimension is infinite
  – Because it can achieve arbitrary decision boundary, i.e., shattering any training set of any size
- Training and LOO error for k-nn
  – Training: zero (some gave LOO error here)
  – LOO for 3-nn: x   x   x   0   0   0   0 x
- PAC learnable: two parts, ln|H| is polynomial, and finding a consistent h in H can be done in polynomial