Announcements

HW 2 due Oct 13
Zillow Prize

k-NN (cont'd)

Recap: which distance function:

How to select k (in k-NN)

small k: noisy large: blurs local effect

select k using validation

portion of misclassified points

n training points
m test points

$$\hat{R}_{test}(f_k) = \frac{1}{m} \sum_{i=n+1}^{n+m} 1\{f_k(x_i) \neq y_i\}$$

Choose k to minimize training data

If we have only training data

- hold out method: randomly assign data points to two sets of size n (training) and m (test).

- k-fold cross validation: randomly partition into k equal sized samples. Use a single sample (partition) as a validation set to test the model and k-

remaining to train the model. Repeat k times with each of the k sub-samples used exactly once as the validation data. Average k results and produce the error estimate.

D_1 validation D_2, D_3, D_4 as training

D_2 D_1, D_3, D_4

⋮

non parametric models

e.g. k-NN

parameters grow with the amount of data.

+ no assumption on dat

- slower, need more data

+ flexible

Other ex: Decision Trees, SVM

parametric models

e.g. Bayes Classifier

parameters is fixed.

e.g. $N(\mu, \Sigma)$

- model may not fit to real data

+ faster to train / test

+ typically requires less data

Other ex: LDA, Logistic Reg, Neural net.

Plug-in Method: Obtained by estimating the terms appearing in the formula for Bayes Classifier.

$$f^*(x) = \arg \max_k \pi_k \cdot g_k(x)$$

$$= \arg\max_k \eta_k(x)$$

$\pi_k = P(Y=k)$ prior class prob.

$g_k(x)$: class conditional pmf/pdf of $X|Y=k$

$\eta_k = P(Y=k|X=x)$ posterior class prob.

Linear Discriminant Analysis (LDA) (Bayesian)

Suppose we have training data

$$(x_1, y_1), \cdots, (x_n, y_n)$$

In LDA we assume $X|Y=k \sim N(\mu_k, \Sigma)$ i.e

$$g_k(x) = \phi(x; \mu, \Sigma)$$

$$= \frac{1}{(2\pi)^{d/2}\sqrt{|\Sigma|}} \exp\left(-\frac{1}{2}(x-\mu_k)^T \Sigma^{-1}(x-\mu_k)\right)$$

LDA is the classifier obtained by plugging in estimates of μ_k, Σ.