SML Week 3

C. Andy Tsao

Dept. of Applied Math
National Dong Hwa University

September 28, 2010

Outline

Bayes classi er for Discrete Scenario

- WLOG, assume $Y = f1; ; Kg$

$EPE(G) = E_{Y;X} L(;G(G;G(E = E_X E$

Estimation of $E(Y|x)$ through KNN or regression can fail

- Curse of dimensionality: KNN includes points afar leads to large error
- If special structure is known, further reduction in bias and variance is possible.

Prediction Problem

- Statistical Model: Assumption on $P_{Y;X}$ (or , say Y (X) +
- Supervised learning


Functional approxima

$E(YjX)$ linear in $X_1, \quad , X_p$

- Simple: easier computation, interpretation and communication
- Readily generalizable: transformation on Y and X, combination of X's'
- Conceptual Framework for more general problems, for example, nonlinear problems.

De nition

$(Y_i; z_i)^{ni}$

How exible is LR?

Assume $_i$ $_{iid}$ N ^{2}d 51

- Y_i = $_0$ + $_1X_i$ + $_i$
- Y_i = $_0$ + $_1X_{1i}X_{2i}$ + $_i$
- *sin*

How exible is LR?

Assume $_i$ $_{iid}$ $N(0; \ ^2)$

- $Y_i = \ _0 + \ _1 X_i + \ _i$
- $Y_i = \ _0 + \ _1 X_{1i} X_{2i} + \ _i$
- $sin(Y_i) = exp(\ _0 + \ _1 X_i) + \ _i$

Your turn

How exible is LR?

Assume $_i$ $_{iid}$ $N(0; \ ^2)$

- $Y_i = \ _0 + \ _1$

Estimation of LR

$Y = X \quad +$

- Solve $\quad$ st $Q(\ ) = j\ Y \quad X\ \ j^2$ is minimized
- Normal equation: $\quad$ solves $X^t(Y \quad X\ ) = 0$:
- When X^tX is nonsingular, $\hat{} = (X^tX)^{-1}X^tY$.
- Geometric Interpretation: $\hat{Y} = X(X^tX)^{-1}X^tY$ is the projection of Y onto the column space of the design matrix X.

Inference: HT and CI

- $\hat{\quad} \quad N(\ ; (X^t X)^{-1} \ ^2)$
- $\hat{}^2 = j \ Y \quad \hat{Y} j^2 = (n \quad p \quad 1):$
- $(n \quad p \quad 1) \ \hat{}^2 \qquad ^2 \ \ ^2_{n-p-1}$

$Y_i = x_i \quad + \quad _i$ (No intercept)

$Y = X \quad + \quad$ where $X = (x_1; \quad ; x_n)^t$

$\hat{} = (X^t X)^{-1} X^t Y = \frac{\sum_n x_i y_i}{\sum_1 x_i^2};$

$r_i = y_i \quad x_i \hat{}.$

In inner product with $< x; y >= \sum_i x_i y_i$

$\hat{} = \frac{<x;y>}{<x;x>}; \quad r = Y \quad X \hat{}:$

Multiple Linear Regression w/ orthogonal x's

- $Y = X + $ where $X = (X_1; ; X_n)^t$
- $\hat{} = (X^t X)^{-1} X^t Y = \frac{\sum_{1}^{n} x_i y_i}{\sum_i x_i^2};$

  $r_i = y_i \quad x_i \hat{}.$
- $\hat{}_j = \frac{<x_j; y>}{<x_j; x_j>}; \quad r = y \quad X \hat{}:$
- When inputs are orthogonal, they have no e ect on each other parameter estimates in the model.

Succession Orthogonalization: Recap

- $\hat{}_j$ represents the additional contribution of X_j on Y, after X_j has been adjusted b01XTyX1; ; Xj1.
- $\hat{Y} = X\hat{}$ is **the** projection of Y onto column space of X
- Non-unique $\hat{}$. Unique $\hat{Y}$
- Alternative iteration for : Iterative residual tting.

Exercise 1: Write down the algorithm for iterative residual tting and show that the obtained $\hat{}$ also solves the normal equation.

Unsatisfying LSE

$YjX_1; \quad ; X_q$, q large/huge

- Accuracy
  Even if $\hat{} = (X^t X)^{-1} X^t Y$ exists, it may have large variance.
- Interpretation
  Non-uniqueness
- Scienti c Important X might be missing
- Variable selection

Subset Selection

$YjX_1;$ $;X_q$, q large/huge. Want to pick $p(<<q)$ X's out of them.

- What have we learned before?

Shrinking Methods

$$\hat{}^{ridge} = argmin \quad (Y \quad X \ )^t (Y \quad X \ ) + \quad ^t$$

- What does this mean? Alternatives?
- (Ex 2) It can be shown

$$\hat{}^{ridge} = (X^t X + \ I)^{-1} X^t Y :$$