SML Week 2

C. Andy Tsao

Dept. of Applied Math
National Dong Hwa University

September 28, 2010

Outline

Review

SDT to SML problem

KNN, LS and More

Formulation of SML problem

Let $(y_i; x_i)_{i=1}^n \overset{iid}{\sim} P_{Y;X}$ with y's $\in Y$; x's $\in X$.
Objective: Find $F \in$ [illegible] such that
$E_{Y;X} L(Y; F(X$ [illegible]

Statistical Decision Theory: Versions of Expected Losses

[Point Estimation Problem]
Let $X_1;$ $;X_n$ f , for example, pdf of $N(; ^2)$ or pmf of *Bernoulli*( ) Objective: Find $\hat{}$

Statistical Decision Theory: Versions of Expected Losses

[Point Estimation Problem]
Let $X_1; \quad ; X_n \quad f$, for example, pdf of $N(\; ; \; ^2)$ or pmf of $Bernoulli(\;)$ Objective: Find $\hat{}$ which minimizes

- Risk: $R(\; ; \hat{}) = E_{Xj} \; L(\; ; \hat{}(X))$
- Bayes expected loss: $E \; _{jx} L(\; ; \hat{}($

Bayes Procedure

We say is a Bayes procedure wrt i

$$= arg \min_{D} r(\ ; \ ):$$

Complete Class Theorem

- A class C is *complete* if for any decision not in C, there exists a decision $^{\theta'}$ which dominates .
- Under some regularity conditions, the class of Generalized Bayes procedures form a complete class.
- Implication: Search no further. Work with Generalized Bayes

KNN as $E(Yjx)$

Let $T = (Y_i; X_i)_{i=1}^{n}$ be the training data.

- $\hat{f}(x) = Ave(y_i jx_i \ 2 \ N_k(x)$, where \Ave" =average, $N_k(x)$ is the neighborhood containing the k points in T closest to x.
- expectation is approximated by averaging over sample space.
-

KNN as $E(Y|x)$

Let $T = (Y_i; X_i)_{i=1}^n$ be the training data.

- $\hat{f}(x) = Ave(y_i|x_i \ 2 \ N_k(x)$, where \Ave" =average, $N_k(x)$ is the neighborhood containing the k points in T closest to x.
- expectation is approximated by averaging over sample space.
- conditioning at a point is relaxed to conditioning on some region \close" to the target point x.

How good is KNN? Rationale? Search is over?

- Sample size usually small
- As p

LS as $E(Y|x)$

- $f(x)$ x^T
- Plug this f into (1), can be solved
  $= [E(XX^T)]^{-1}E(XY)$

KNN vs. LS

- LS assumes $f(x)$ globally by a linear function
- KNN assumes $f(x)$ locally by a const function

Homework 1 (due 101012):
Ex 1: Follow $x2.5$ in HTF and use R, reproduce Fig 2.7 and Fig 2.9.
You may need to install contributed R packages such as kknn.
Ex 2{4: Ex 2.1, Ex 2.5 and Ex 2.6 in HTF.