IDS 575

1. $P_{XY}(X = x, Y = y)$

$P_{X_1 X_2 Y}(X_1 = x_1, X_2 = x_2, Y = y)$

X_1	X_2	Y	$P(\ )$
0	0	0	1/8
0	0	1	1/8
⋮			
1	1	1	1/8

X : p dim

Y : 1 dim

$X = [\ ]_{N \times p}$ $\quad Y = [\ ]_{N \times 1}$

Expected (squared) prediction error (EPE)

$$= E_{XY}\left[\left(Y - f(X)\right)^2\right]$$

obj

$$= \sum_{x} \sum_{y} P(X=x, Y=y) \cdot (y - f(x))^2$$

loss function

$$\rightarrow (Y - f(X))^2$$

$$\sum_{i=1}^{N} (y_i - f(x_i))^2$$

Goal: find the best function 'f'

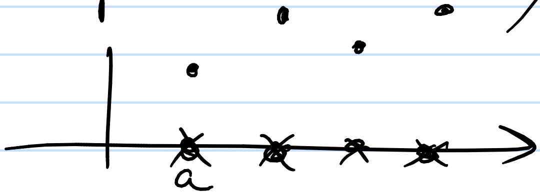
$$= \sum_{x} \sum_{y} P(Y=y \mid X=x) \cdot P(X=x) \, (y - f(x))^2$$

$$= \underbrace{\sum_x P(X=x)}_{E_X} \underbrace{\left[\sum_y P(Y=y|X=x)\,(y-f(x))^2\right]}_{E_{Y|X}}$$

$$= \min_f E_X\left[\underbrace{E_{Y|X}\left[(y-f(x))^2\right]}_{\text{depends on } x}\right]$$

X	Y	P( )
a	.	.
b	.	.
c	.	.
d	.	.

X is 1 dim



$$E_x\left[\underbrace{E_{Y|X}\left[(Y-\underline{f}(x))^2\right]}\right]$$

$$\frac{d}{dz} z^2 = 2z$$

$$\frac{d}{dz}(y-z)^2 = -2(y-z)$$

$$\min_z \underbrace{\sum_y P(Y=y \mid \underline{X=x}) \cdot (y-z)^2}$$

Set derivative equal to 0

$$\sum_y P(Y=y \mid X=x)(y-z) = 0$$

$$\Rightarrow \underbrace{\sum_y P(Y=y \mid X=x) \cdot y} = z \underbrace{\sum_y P(Y=y \mid X=x)}_{1}$$

$$\underline{\underline{E[Y \mid X = x']}} = f(x')$$

$$f(x) = \underline{\underline{E[Y \mid X = x]}}$$ ← "regression function"

① k-n-n $\qquad$ $k = 3$

$$\underbrace{\begin{bmatrix} 0 & 1 \\ 1 & 2 \\ 5 & 4 \\ 3 & 3 \end{bmatrix}}_{X} \qquad \underbrace{\begin{bmatrix} 1 \\ 2 \\ 3 \\ 0 \end{bmatrix}}_{Y}$$

$x^{test} = [2, 2]$

3 nearest neighbors

$$\boxed{\begin{array}{l} \text{k-n-n} \\ \text{① 'K' ② \underline{distance}} \end{array}}$$

$$\underbrace{E[Y \mid X = x]}\ \overset{①}{\text{—}}\ \approx \frac{1}{K} \sum_{i \text{ such that } x_i = x} y_i \ \overset{②}{\approx}\ \underbrace{\frac{1}{K} \sum_{i \text{ such that } x_i \approx x} y_i}$$

(K)

$$Y = \begin{cases} 0 & 1/2 \\ 1 & 1/2 \end{cases} \qquad E[Y] = 1/2$$

$$y_1 \ldots y_{100} \qquad E[Y] \approx \frac{1}{N} \sum y_i$$

$\cancel{X}$ G majority vote$(g_i$ ~~of~~ $i : x_i \approx x)$

Linear model $f(x) = \beta^T x = [\beta^1 \ \beta^2] \begin{bmatrix} x^1 \\ x^2 \end{bmatrix} = \beta^1 x^1 + \beta^2 x^2$

$$f(x) = \underbrace{E[Y \mid X = x] = \beta^T x}$$

$$\min_{\beta} EPE = \min_{\beta} E_{XY}\left[(Y - \beta^T X)^2 \right]$$

$$\beta = \left(E_X [X X^T] \right)^{-1} E_{XY}(X \cdot Y)$$

$$\underset{2\times1}{\begin{bmatrix} x^1 \\ x^2 \end{bmatrix}} \underbrace{\begin{bmatrix} x^1 & x^2 \end{bmatrix}}_{1\times2} = \underset{2\times2}{\begin{bmatrix} & \\ & \end{bmatrix}}$$

Data: X, Y $\qquad \hat{\beta} = (\overset{M}{X^T X})^{-1} X^T Y$ $\qquad X = [\ ]_{N \times d} \quad Y = [\ ]_{N \times 1}$

Dist $X\ Y$ $\qquad \beta^{dist} = (\underbrace{E[X X^T]})^{-1} E[XY]$

Q: is M not invertible? $\qquad p \times p$

$$\begin{bmatrix} | & | & | \end{bmatrix} = \underbrace{\begin{bmatrix} 5 & 1 & 6 \\ 2 & 2 & 4 \\ 3 & 1 & 4 \\ 1 & 1 & 2 \end{bmatrix}}_{N \times p}$$

$$\begin{bmatrix} 1 & 5 & 2 \\ 2 & 6 & 4 \\ 3 & 7 & 6 \end{bmatrix}$$

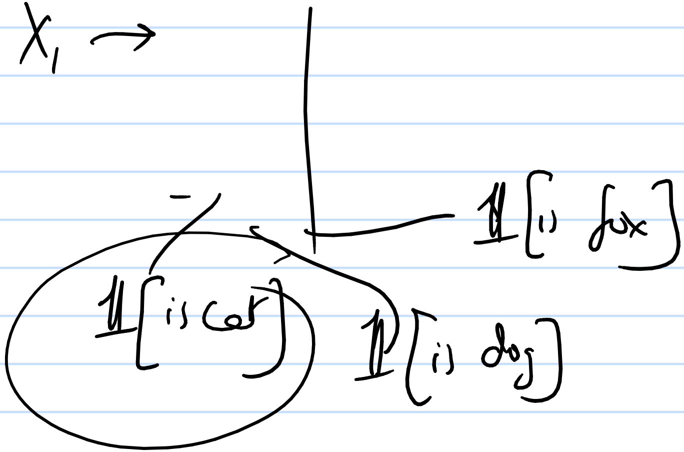
$X_1 = \{Cat, dog, fox\}$

$$X \begin{bmatrix} 1 & Cat \\ 2 & dog \\ 3 & fox \\ 1 & dog \end{bmatrix}$$

$$\sum_{[\ ], y} P\left(X = \begin{bmatrix} x^1 \\ \vdots \\ x^p \end{bmatrix}, Y = y\right) \cdot \begin{bmatrix} x^1 \\ \vdots \\ x^p \end{bmatrix} \cdot y$$

$$X = \begin{bmatrix} 1 & 1 & 0 & 0 & 1 \\ 2 & 0 & 1 & 0 & 1 \\ 3 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 \end{bmatrix}$$

$X_1 \rightarrow$

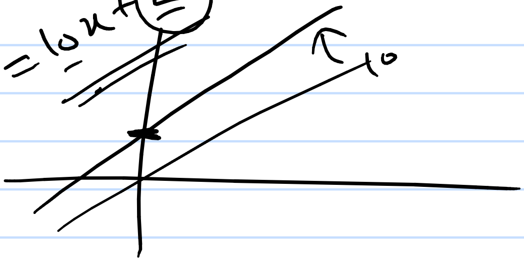


$$f(x) = \beta^T x + C$$

$$= \begin{bmatrix} \beta^1 & \beta^2 & C \end{bmatrix} \begin{bmatrix} x^1 \\ x \\ 1 \end{bmatrix}$$

$$\begin{bmatrix} 10 & 2 \end{bmatrix} \begin{bmatrix} x \\ 1 \end{bmatrix}$$

$$y = 10x + 2$$



② EPE for classification → Bayes classifier

③ Issues with k-nn

④ Relatives of EPE, k-nn & linear models

Scoring for classification.

$$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \quad \begin{bmatrix} 0 & 5 \\ 10 & 0 \end{bmatrix}$$

$$\begin{bmatrix} 0 & 5 & 5 \\ 5 & 0 & 10 \\ 5 & 5 & 0 \end{bmatrix}_{3\times 3} = L \Leftrightarrow L\left(G = \overset{1,2,3}{k},\ \hat{G}(x) = k'\right)$$

$$EPE = E_X\left[\sum_{k=1}^{K} P(G=k|X) \cdot L(G=k, \hat{G}(X))\right]$$

$$\begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}$$

$$\min_{z} \sum_{k=1}^{K} P(G=k|X=x) \cdot L(G=k, z)$$

$$P_{GX}(G=k, X=x)$$

$z = \text{cat}$

$z = \text{dog}$

$z = \text{fox}$

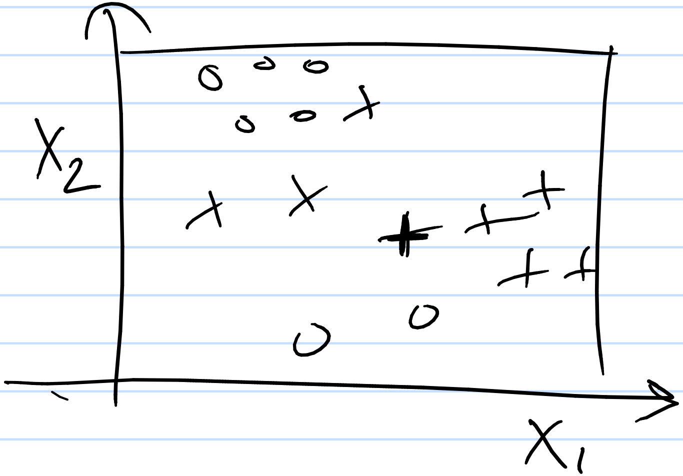
$P(G=\text{cat}|x) = .3$

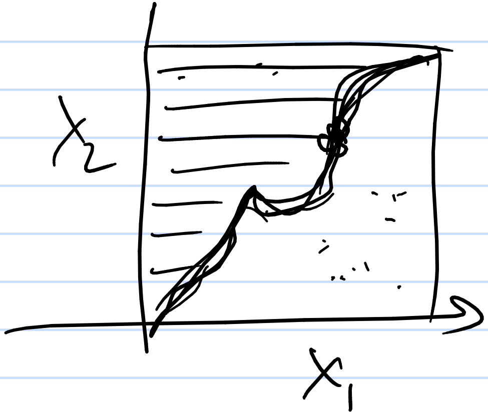
$\text{dog}|x) = .4$

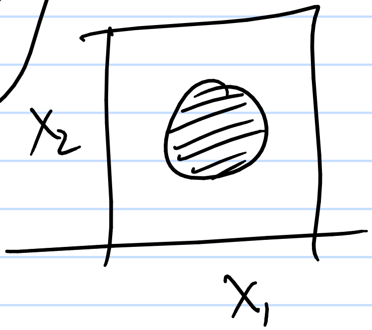
$\text{fox}|x) = .3$

best classifier for $X = x$ is

$$\underset{k=1,2,3}{\arg\max} \; \underbrace{P(G=k \mid X=x)}$$ — Bayes classifier

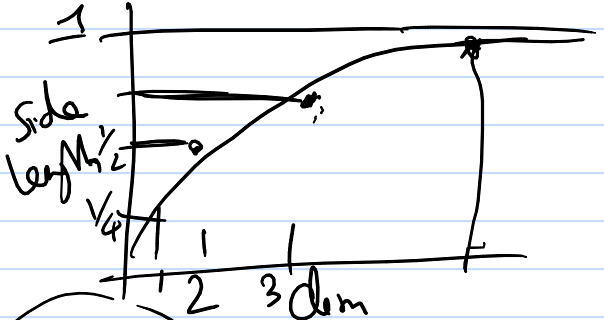






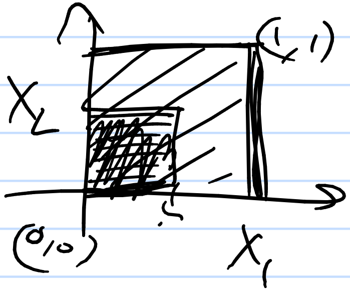
(2) Why is k-nn not enough.

defining "neighbors"



p

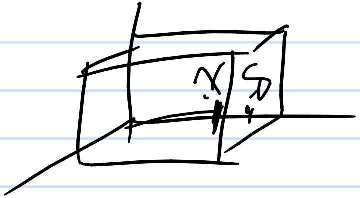
N are finite



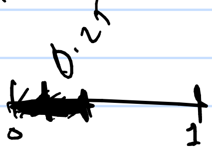
1d $1/4$ $1/4$

2d: region = $1/4$ side length = $1/2$ = $r^{1/2}$

3d: $1/4$ $\left(\frac{1}{4}\right)^{1/3}$



p dim $\left(\frac{1}{4}\right)^{1/p}$



Relatives

Knn $\longrightarrow$ Kernel method

$$\text{pred} = \frac{\sum_{i=1,\dots N} K(x_i, x^{test}) \cdot y_i}{\sum_{i=1,\dots N} K(x_i, x^{test})}$$

linear model $\longrightarrow$?