STAT 391

3/7/2025

Lecture 18

Classifiers Ch 17

Statistical decisions 15

(Probabilistic reasoning)[14]

Classification

- Logistic Regression ✓
  linear, discriminative

{ Discriminative
{ Generative

- Nearest Neighbors

- Generative
  - Fisher discriminant
  - "language (

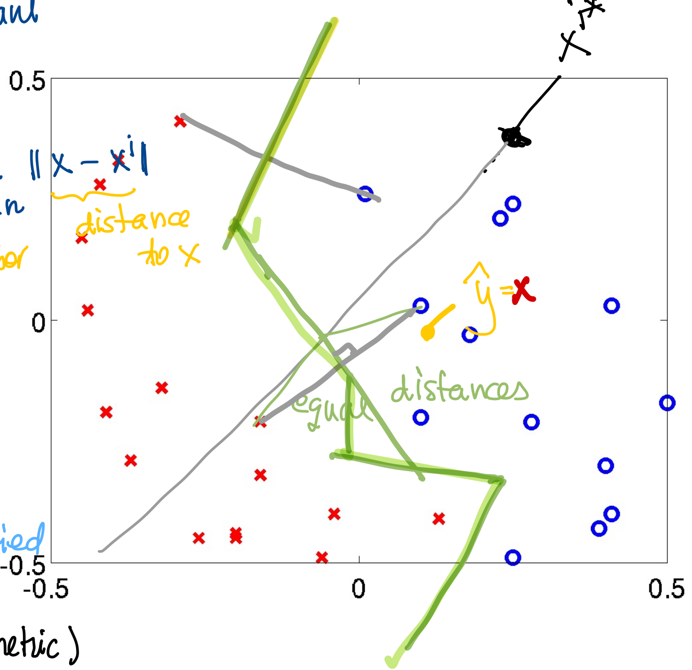
NN classifier $\mathcal{D} = \{(x^{1:n}, y^{1:n})\}$

given x: 1. find $x^{i^*} = \underset{i=1:n}{\operatorname{argmin}} \underbrace{\| x - x^i \|}_{\text{distance to } x}$

$\uparrow$ nearest neighbor

2. $\hat{y}(x) = y^{i^*}$

$\hat{y} = o$

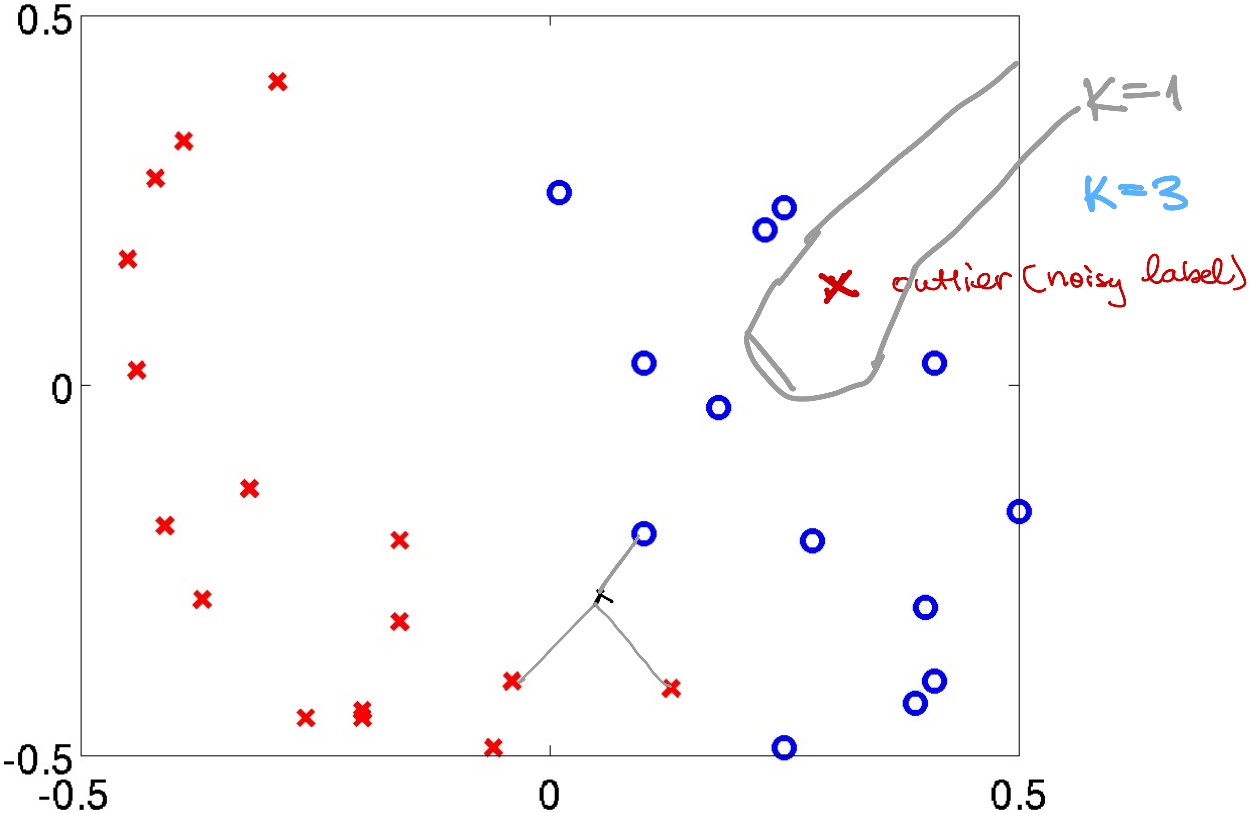


- D. boundary = piecewise linear
- all x^i are correctly classified (Bias = 0)
- Var large (non-parametric)

K- nearest neighbors (K-nn)

nn1(x) = o K=1 o

nn2,3(x) = × K=3 ×



Generative classifiers

$y \in \{0, 1\}$

OR $y \in \{1, 2, \dots C\}$ multiway

Train $\mathcal{D} = \{(x^{1:n}, y^{1:n})\}$

$\mathcal{F}$ = model class

· for $c = 1, 2, \dots C$ each class

1. Estimate $f_c \in \mathcal{F}$ from $\mathcal{D}_c = \{x^i \mid y^i = c\}$

generative models

learned $P_{X|Y=c}$

[But want $P_{Y|X}$!]

2. Estimate $\pi_c = \frac{n_c}{n} = \frac{|\mathcal{D}_c|}{|\mathcal{D}|}$ = Prob of class c

$\sum_{c=1}^{C} \pi_c = 1$

Ex 1 digits $c = \{0, \dots 9\}$

f_0 = density estimator on all "0" digits

f_2 = — " — on all "2" digits

$f_3 = \dots$

Ex 2 Languages

train on English, Spanish, German text

$P^E = (\theta^E_{a,b,c \dots})$

$P^S = (\theta^S_{a,b,\dots})$

Prediction (classification) with gen. model

given x: calculate $f_c(x)$ likelihood of $x \mid Y=c$

new

$\hat{y} = \arg\max_x f_c(x)$

$Pr[y = c \mid x] =$

Max Likelihood ("Likelihood Ratio")

Bayesian

MAP Max A-Posteriori

$$\Pr[y=c \mid x] = \frac{\overset{\text{prior}}{\pi_c = P(c)}\ \overset{\text{likelihood}}{f_c(x)}}{\underbrace{\sum_{c'} P(c')\, f_{c'}(x)}_{\text{normalization}}}$$

MAP $\quad \hat{y}(x) = \underset{c}{\operatorname{argmax}}\ P[y=c \mid x] = \underset{c}{\operatorname{argmax}}\ \pi_c \cdot f_c(x)$

Fisher discriminant

$y \in \{0,1\}$ $\qquad x \in \mathbb{R}^d \qquad [d=1]$ or $d=2$

$f_0 = N(\mu_0, \sigma^2 I_d)$

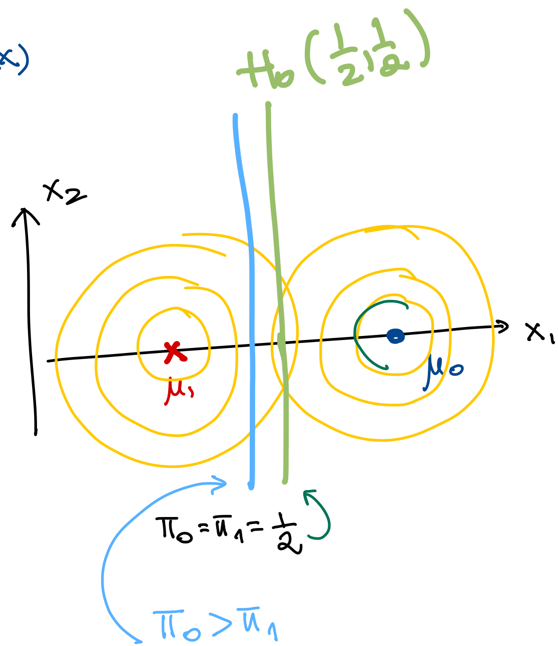
$f_1 = N(\mu_1, \sigma^2 I_d)$

same covariance

π_0, π_1 = any values

Ex: $f(x) = \dfrac{\pi_0 f_0(x)}{\pi_1 f_1(x)} \gtrless 1 \qquad \begin{matrix} \hat{y}=0 \\ \hat{y}=1 \end{matrix}$

$\{x : f(x) = 0\}$ = hyperplane, OR $f(x)$ linear in x



Statistical decisions

Given x = input $\in \mathcal{X} = \mathbb{R}^d$

a = action $\in \mathcal{A}$ → ex:

1. Classification : a = choose a class
2. Sequential decisions
3. —"— games

$\text{Cost}(y)$, $\text{Cost}(y, x, a)$

↑ outcome

Ex 1 How many servers ?

x = demand = # requests / min

a = # servers = $\{1, 2, \dots M\}$

outcomes $y = \begin{cases} x \le a & \text{all requests served} \\ x > a & x-a \text{ requests dropped} \end{cases}$

~~costs~~ gains $= \begin{cases} \text{served} & 1 \\ \text{dropped} & -5 \ * \\ \text{running server} & -0.5 \end{cases}$

P_x = Poisson (λ)

↑ known

Max Expected Gain

Min —"— Loss

$$\underset{a}{\text{argmax}} \underset{P_x}{\mathbb{E}}[\text{gain}(x)] = E_{P_x}\left[1 \cdot \min(x, a) - 0.5 \cdot a - 5 \cdot \max(x-a, 0)\right] =$$

$$= 1 \cdot \sum_{x \le a} P(x) \cdot x - 0.5a - 5 \cdot \sum_{x > a} (x-a) P(x)$$