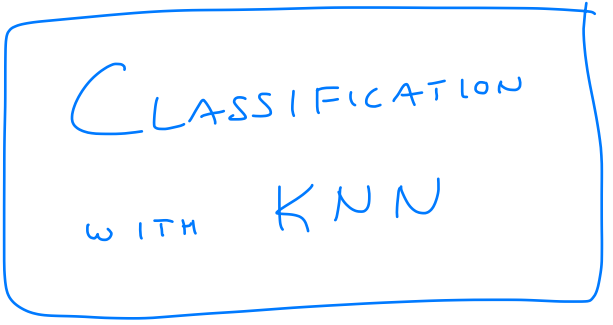
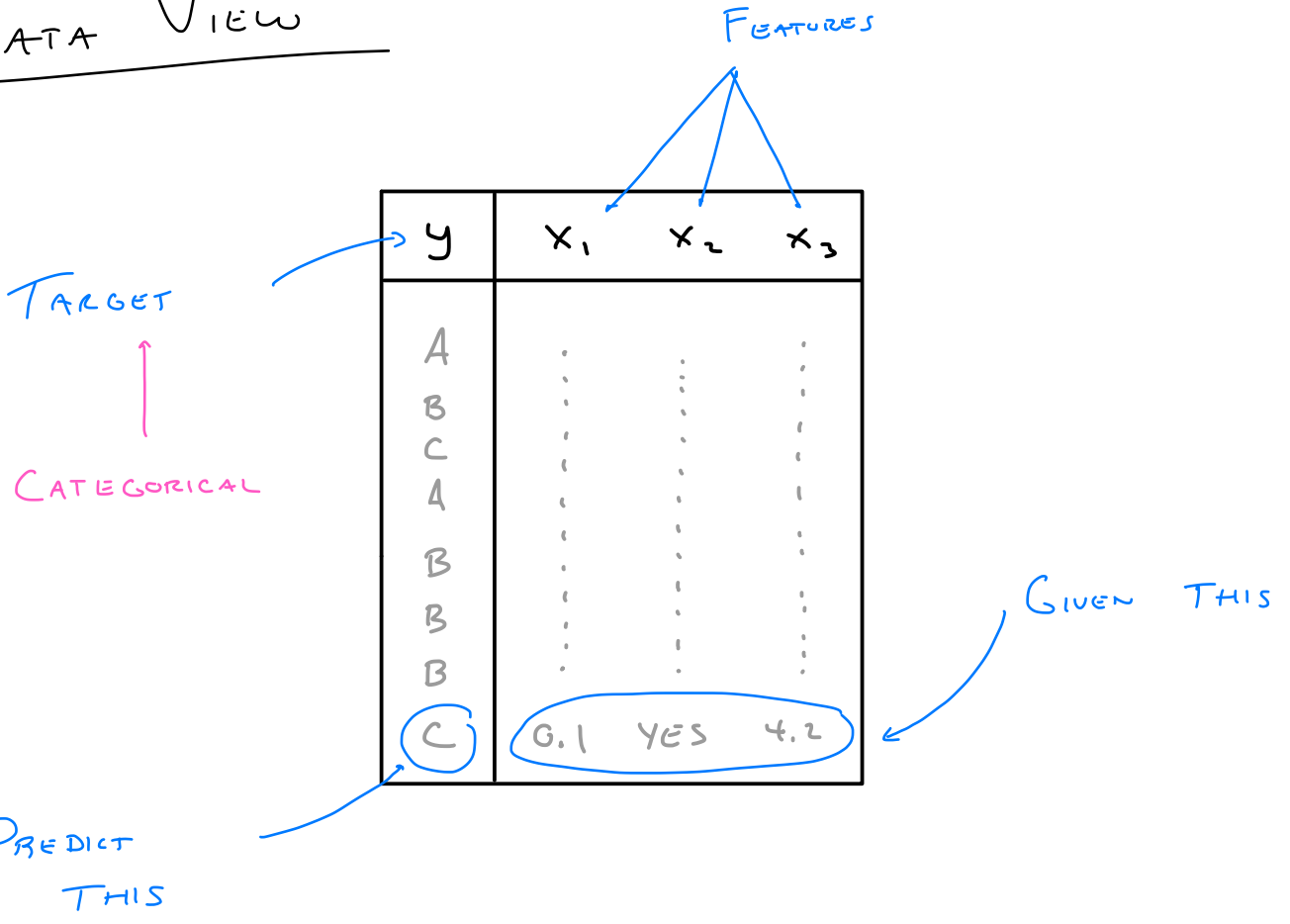
CS 307



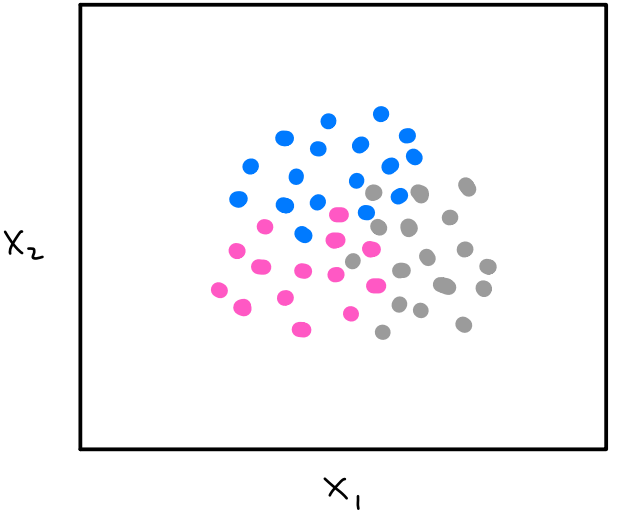
CLASSIFICATION

AN INTRODUCTION

Data View


Features
y
x_1
x_2
x_3
Target
Categorical
A
B
C
A
B
B
B
C
6.1
YES
4.2
Given This
Predict This

Visual Data View



Probability View

$$(X, Y) \in \mathbb{R}^p \times \{1, 2, \ldots K\}$$

p Features

K Categories

Features

Response

Find a classifier $C(x)$ that minimizes

$$P[C(x) \neq Y]$$

← Probability of misclassification

where $C : \mathbb{R}^p \longrightarrow \{1, 2, 3, \ldots K\}$

Input features

Output category

BAYES CLASSIFIER

← MINIMIZES PROBABILITY OF MISCLASSIFICATION

$$C^B(x) \triangleq \underset{k \in \{1,\ldots,K\}}{\text{ARGMAX}} \; P[Y = k \mid X = x]$$

GIVEN FEATURE VECTOR x, CLASSIFY OBSERVATION AS THE CATEGORY WITH THE HIGHEST PROBABILITY

DUH?

EXAMPLE

$$C^B(x=0) = ?$$

$$P[x=0 \cap Y=A] \over P[x=0]$$

Y \ X	0	1	
A	0.1	0.1	0.2
B	0.2	0.1	0.3
C	0.1	0.4	0.5
	0.4	0.6	

JOINT DISTRIBUTION OF (X,Y)

MARGINAL DISTRIBUTION OF X

$$P[Y \mid X=0] = \begin{cases} 0.25 & y=A \\ 0.50 & y=B \\ 0.25 & y=C \end{cases}$$

CONDITIONAL PROBABILITY OF $Y \mid X=0$

$$C^B(x=0) = B$$

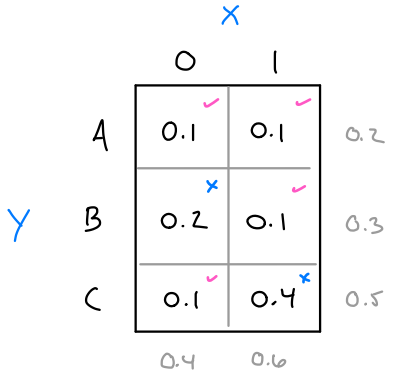
$$C^B(x=1) = C$$

Bayes Error

← average misclassification using Bayes classifier

$$1 - E_x \left[\max_k P[Y = k \mid X = x] \right]$$

"irreducible error"

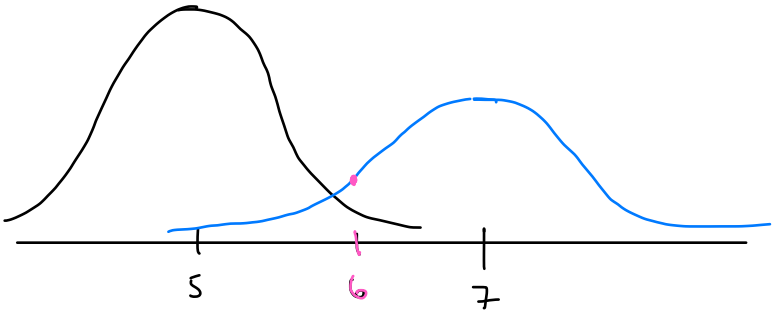


$$= 1 - \left[P[Y = B \mid X = 0] P[X = 0] + P[Y = C \mid X = 1] P[X = 1] \right]$$

$$= 1 - \left[\left(\frac{0.2}{0.4} \right)(0.4) + \left(\frac{0.4}{0.6} \right)(0.6) \right]$$

$$= 1 - [0.2 + 0.4] = 0.4$$

EXAMPLE



$$X \mid Y=0 \sim N(\mu=5, \sigma=1) \qquad f_0(x)$$

$$X \mid Y=1 \sim N(\mu=7, \sigma=2) \qquad f_1(x)$$

$$\pi_0 = P[Y=0] = 0.6$$

$$\pi_1 = P[Y=1] = 0.4$$

$$C^B(x=6) = ?$$

CALCULATE $P[Y=0 \mid X=6]$

$$= \frac{\pi_0 f_0(6)}{\pi_0 f_0(6) + \pi_1 f_1(6)} = \text{----}$$

TO SCIPY!

$$P[Y=0 \mid X=6] = \frac{\pi_0 f_0(6)}{\pi_0 f_0(6) + \pi_1 f_1(6)}$$

$$P[Y=1 \mid X=6] = \frac{\pi_1 f_1(6)}{\pi_0 f_0(6) + \pi_1 f_1(6)}$$

SAME DENOMINATOR

ONLY NEED NUMERATOR FOR CLASSIFICATION

In Practice

Don't know $P[Y = k \mid X = x]$!!!

Estimate it !!!

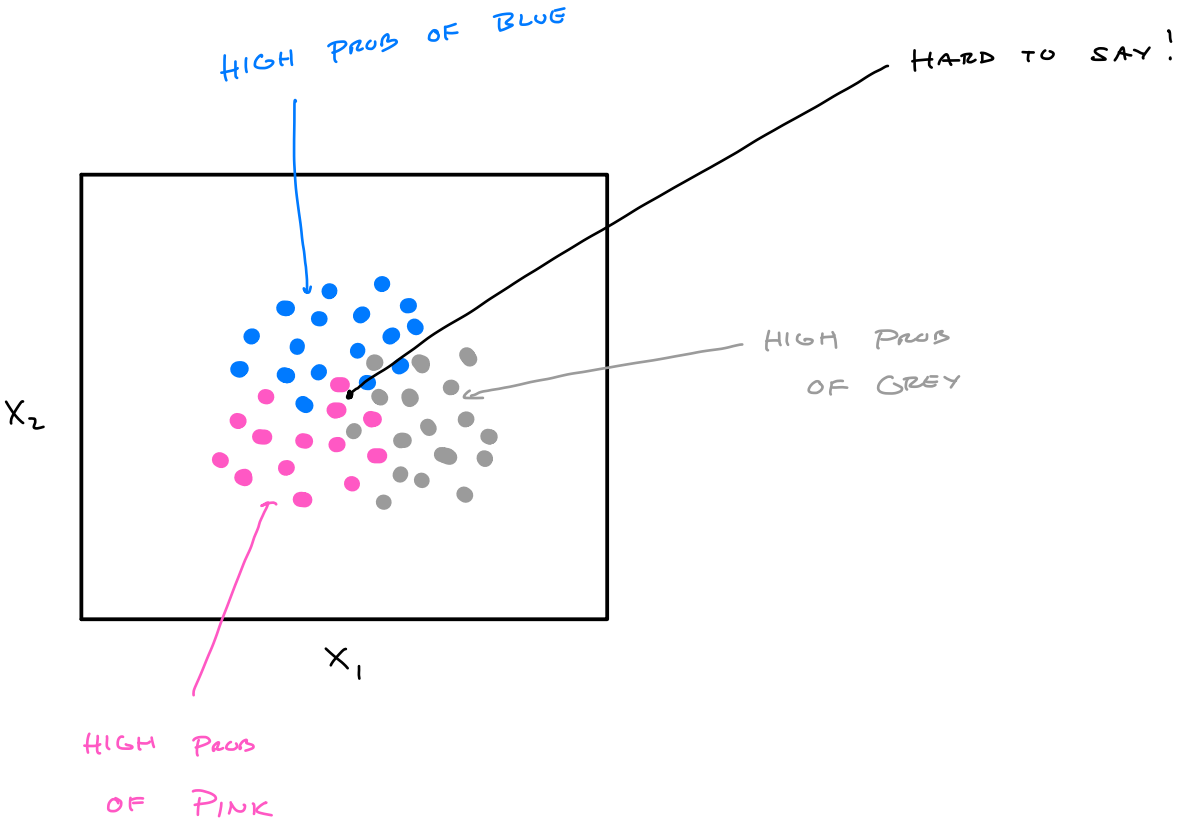
Learned classifier

$$\hat{C}(x) = \underset{k}{\text{argmax}}\ \hat{P}[Y = k \mid X = x]$$

$\hat{P}[Y = k \mid X = x]$: estimate of conditional probability

A "guess" for $C^B(x)$

How ?


HIGH PROB OF BLUE
HARD TO SAY!
HIGH PROB OF GREY
X_2
X_1
HIGH PROB OF PINK

Estimating Conditional Probabilities

KNN w/ sklearn.neighbors.KNeighborsClassifier

Trees w/ sklearn.tree.DecisionTreeClassifier

Linear models w/ sklearn.linear_model.LogisticRegression

} Familiar Interfaces

clf.fit

clf.predict ← make classifications

clf.predict_proba ← estimate conditional probabilities

METRICS

WOULD LIKE $P\left[C(x) \neq Y\right]$

SETTLE FOR $\frac{1}{n}\sum_{i=1}^{n} I\left(C(x_i) \neq y_i\right)$

MISCLASSIFICATION

$$I\left(C(x_i) \neq y_i\right) = \begin{cases} 1 & C(x_i) \neq y_i \\ 0 & C(x_i) = y_i \end{cases}$$

$\frac{1}{n}\sum_{i=1}^{n} I\left(C(x_i) = y_i\right)$ ← ACCURACY

$C(x)$ PLACEHOLDER

$\hat{C}(x)$ LEARNED

$C^{*}(x)$ BAYES

KNN CLASSIFICATION

- MOSTLY THE SAME AS REGRESSION
- FOCUS ON ESTIMATING

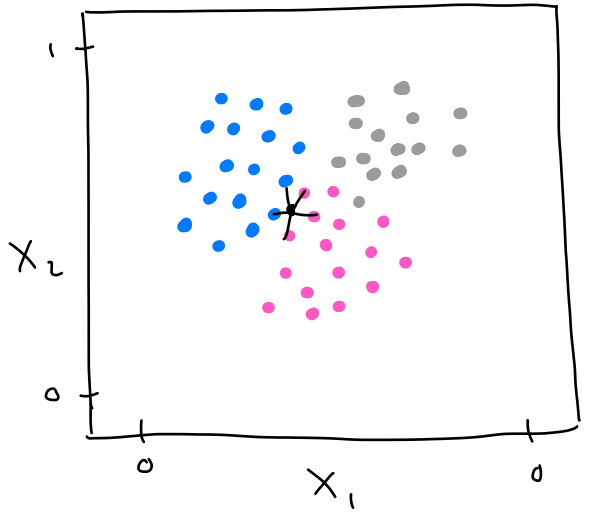
$$P[Y = g \mid X = x]$$

KNN Classification

How to estimate?

$$\hat{P}\left[Y = g \mid X = x\right] = \frac{1}{k} \sum_{\{i : x_i \in \mathcal{N}_k(x, D)\}} I(y_i = g)$$

$K = 5$

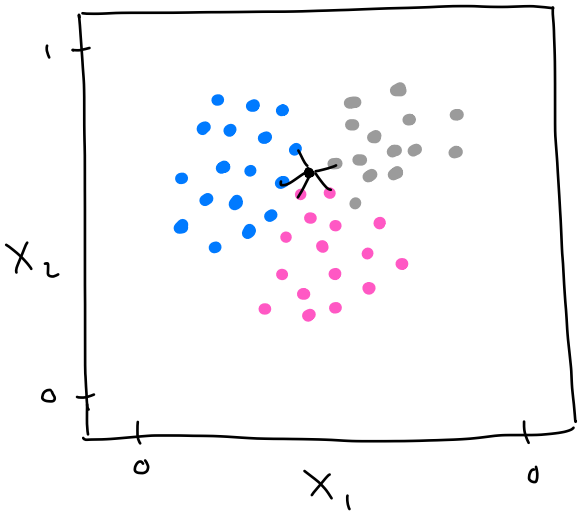


$\hat{p}[Y = \bullet \mid X = \bullet] = 0/5$

$\hat{p}[Y = \bullet \mid X = \bullet] = 2/5$

$\hat{p}[Y = \bullet \mid X = \bullet] = 3/5$

$k = 5$



$$\hat{p}[Y = \bullet \mid X = \bullet] = 1/5$$

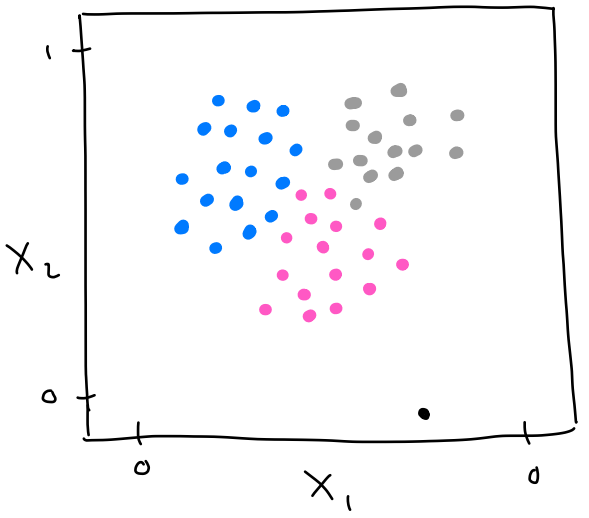
$$\hat{p}[Y = \bullet \mid X = \bullet] = 2/5$$

$$\hat{p}[Y = \bullet \mid X = \bullet] = 2/5$$

How to classify?

↳ Depends who wrote code!

$K = 5$



$$\hat{P}[Y = \bullet \mid X = \bullet] = 0/5$$

$$\hat{P}[Y = \bullet \mid X = \bullet] = 0/5$$

$$\hat{P}[Y = \bullet \mid X = \bullet] = 5/5$$

BUT SHOULD YOU CLASSIFY?