KNN & NB

Naïve Bayes Classifier

greatlearning
Learning for Life

Naive Bayes Classification

- Will my flight be on time? It is Sunny, Hot, Normal Humidity, and not Windy!
- Data from the last several times we took this flight

OUTLOOK	TEMPERATURE	HUMIDITY	WINDY	Flight On Time
Rainy	Hot	High	FALSE	No
Rainy	Hot	High	TRUE	No
Overcast	Hot	High	FALSE	Yes
Sunny	Mild	High	FALSE	Yes
Sunny	Cool	Normal	FALSE	Yes
Sunny	Cool	Normal	TRUE	No
Overcast	Cool	Normal	TRUE	Yes
Rainy	Mild	High	FALSE	No
Rainy	Cool	Normal	FALSE	Yes
Sunny	Mild	Normal	FALSE	Yes
Rainy	Mild	Normal	TRUE	Yes
Overcast	Mild	High	TRUE	Yes
Overcast	Hot	Normal	FALSE	Yes
Sunny	Mild	High	TRUE	No

$A \rightarrow$ event

$P(A) = 0.5$

$P(\bar{A}) = 1 - P(A)$

$A \rightarrow$ Roll a 1

$B \rightarrow$ Roll an odd # number

$P(A) = 1/6$

$P(B) = 3/6 = 1/2$

Cond. prob

$P(A) = 1/6$

		$P(A \mid B)$
1	✓	✓
2	×	
3	×	×
4	×	
5	×	×
6	×	

$$P(A \mid B) = 1/3 = \frac{P(A \text{ and } B)}{P(B)}$$

$$\frac{(1/6)}{(1/2)}$$

greatlearning
Learning for Life

4) A → Roll ① on dice $\boxed{1}$

B → Roll a ① on dice $\boxed{2}$

$P(A \text{ and } B) = \frac{1}{36} = P(A)\,P(B)$

$P(A \mid B) = \frac{1}{6} \Leftarrow P(A \mid B) = \frac{P(A \text{ and } B)}{P(B)}$

$P(B \mid A) = \frac{1}{6}$

$= \frac{P(A)\,\cancel{P(B)}}{\cancel{P(B)}}$

$\boxed{P(A \mid B) = P(A)}$

$\boxed{P(A \text{ and } B) = P(A)\,P(B)}$

$$P(A \text{ and } B) = P(A|B)\,P(B)$$

$$P(B \text{ and } A) = P(B|A)\,P(A)$$

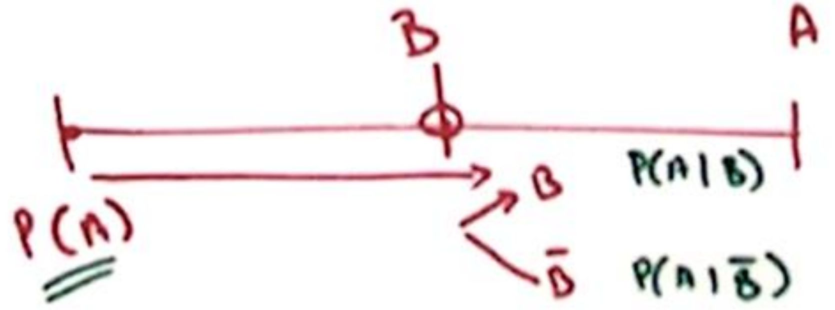
Prior

Bayes thm.

$$\boxed{P(A|B)\,\cancel{P(B)} = \frac{P(B|A)\,P(A)}{P(B)}}$$

Posterior

$$P(A|\bar{B}) = \frac{P(\bar{B}|A)\,P(A)}{P(\bar{B})}$$



- You are planning an outdoor event tomorrow. When it actually rains, the weatherman correctly forecasts rain 90% of the time. When it doesn't rain, he incorrectly forecasts rain 10% of the time. Historically it has rained only 5 days each year. Unfortunately, the weatherman has predicted rain for tomorrow. What is the probability that it will rain tomorrow?

✓ $F \Rightarrow$ Forecast Rain

$\bar{F} \Rightarrow$ Forecast no Rain

P(R)

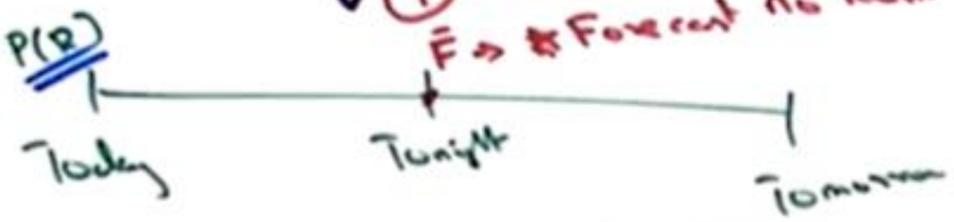
Today — Tonight — Tomorrow

R - Rain tomorrow $\boxed{P(R) = \frac{5}{365}} = 0.014$

$$P(R \mid F) = \frac{P(F \mid R)\, P(R)}{P(F)}$$

$F \Rightarrow$ Forecast Rain

$\bar{F} \Rightarrow$ Forecast No Rain



R - Rain tomorrow $\boxed{P(R) = \frac{5}{365}} = (0.014)$

$$P(R \mid F) = \frac{\overset{(0.9)}{P(F \mid R)}\, P(R)}{P(F)} = \boxed{0.11}$$

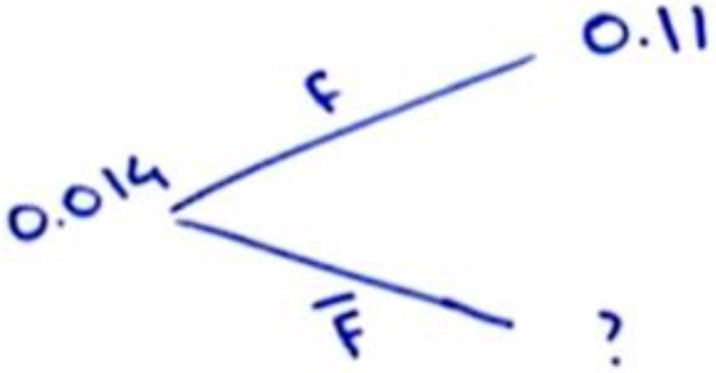
$$P(F) = P(F \text{ and } R) + P(F \text{ and } \bar{R})$$

$$= P(F \mid R)\, P(R) + P(F \mid \bar{R})\, P(\bar{R})$$

$$= (0.9)(0.014) + (0.1)\left(\frac{360}{365}\right)$$

$P(R \mid F) = 0.11$

$P(R \mid \bar{F}) =$



Naive Bayes Classification

greatlearning
Learning for Life

- Will my flight be on time? It is Sunny, Hot, Normal Humidity, and not Windy!
- Data from the last several times we took this flight

OUTLOOK	TEMPERATURE	HUMIDITY	WINDY	Flight On Time
Rainy	Hot	High	FALSE	No
Rainy	Hot	High	TRUE	No
Overcast	Hot	High	FALSE	Yes
Sunny	Mild	High	FALSE	Yes
Sunny	Cool	Normal	FALSE	Yes
Sunny	Cool	Normal	TRUE	No
Overcast	Cool	Normal	TRUE	Yes
Rainy	Mild	High	FALSE	No
Rainy	Cool	Normal	FALSE	Yes
Sunny	Mild	Normal	FALSE	Yes
Rainy	Mild	Normal	TRUE	Yes
Overcast	Mild	High	TRUE	Yes
Overcast	Hot	Normal	FALSE	Yes
Sunny	Mild	High	TRUE	No
Sunny	Hot	Normal	False	?

Bayes' Rule Continued

- In words, Bayes' rule says that the posterior is the likelihood times the prior, divided by a sum of likelihoods times priors.
- The denominator in Bayes' rule is the probability P(B).

$$\boxed{\text{posterior probability} = \frac{\text{conditional probability} \cdot \text{prior probability}}{\text{evidence}}}$$

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

$$P(\overbrace{\text{Flight on time}}^{A} \mid \overbrace{\text{Sunny, Hot, Nor, False}}^{B})$$

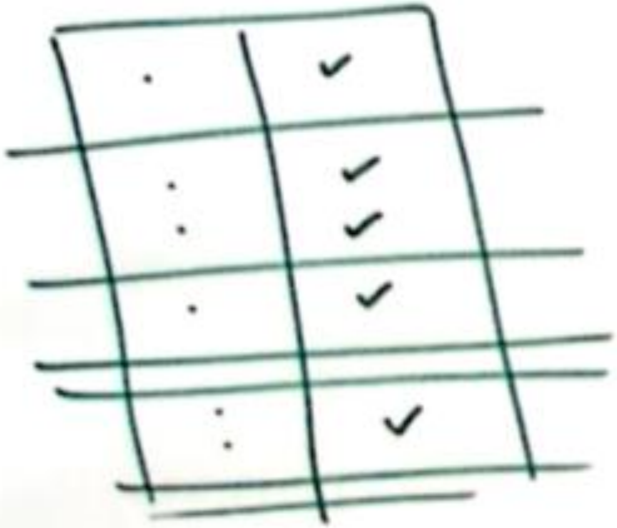
$$\frac{P(S,H,N,F \mid T)\, P(T)}{P(S,H,N,F)}$$

$$P(T) = 9/14$$

$$P(S, H, N, F \mid T) = \frac{P(S \mid T)}{T} \; \frac{P(H \mid T)}{T} \; \frac{P(N \mid T)}{T} \; \frac{P(F \mid T)}{T}$$

$$= \left(\frac{2}{9} \cdot \frac{2}{9} \cdot \frac{6}{9} \times \frac{6}{9} \right)$$

T?



$$\boxed{P(T \mid S,H,N,F)} \Rightarrow 67\%$$

Bayes

$$\frac{\boxed{P(S,H,N,F \mid T)\,P(T)}}{P(S,H,N,F)} = \frac{0.0141}{P(S,H,N,F)} = \frac{0.0141}{0.0141+0.0068} = \boxed{0.67}$$

$$P(S,H,N,F) = P(S,H,N,F \text{ and } T) + P(S,H,N,F \text{ and } \bar{T})$$

$$= 0.0141 + P(S,H,N,F \mid \bar{T})\,P(\bar{T})$$

$$P(S \mid \bar{T})\,P(H \mid \bar{T})\,P(N \mid \bar{T})\,P(F \mid \bar{T})\quad P(\bar{T})$$

$$\underbrace{3/5 \cdot 2/5 \cdot 1/5 \cdot 2/5 \cdot 5/14}_{0.0068}$$

Naïve Bayes Classifiers

- Probabilistic models based on Bayes' theorem.
- It is called "naive" due to the assumption that the features in the dataset are mutually independent
- In real world, the independence assumption is often violated, but naïve Bayes classifiers still tend to perform very well
- Idea is to factor all available evidence in form of predictors into the naïve Bayes rule to obtain more accurate probability for class prediction
- It estimates conditional probability which is the probability that something will happen, given that something else has already occurred. For e.g. the given mail is likely a spam given appearance of words such as "prize"
- fBeing relatively robust, easy to implement, fast, and accurate, naive Bayes classifiers are used in many different fields

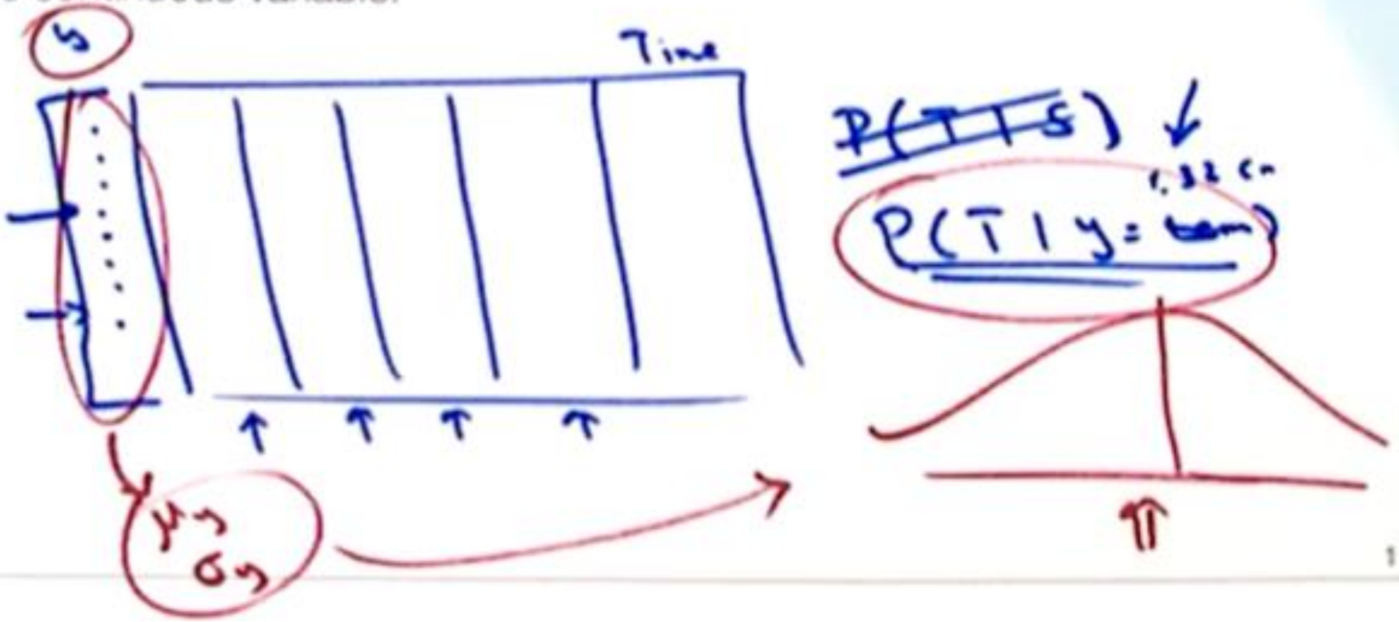
Naïve Bayes Classifiers - Pros and Cons

greatlearning
Learning for Life

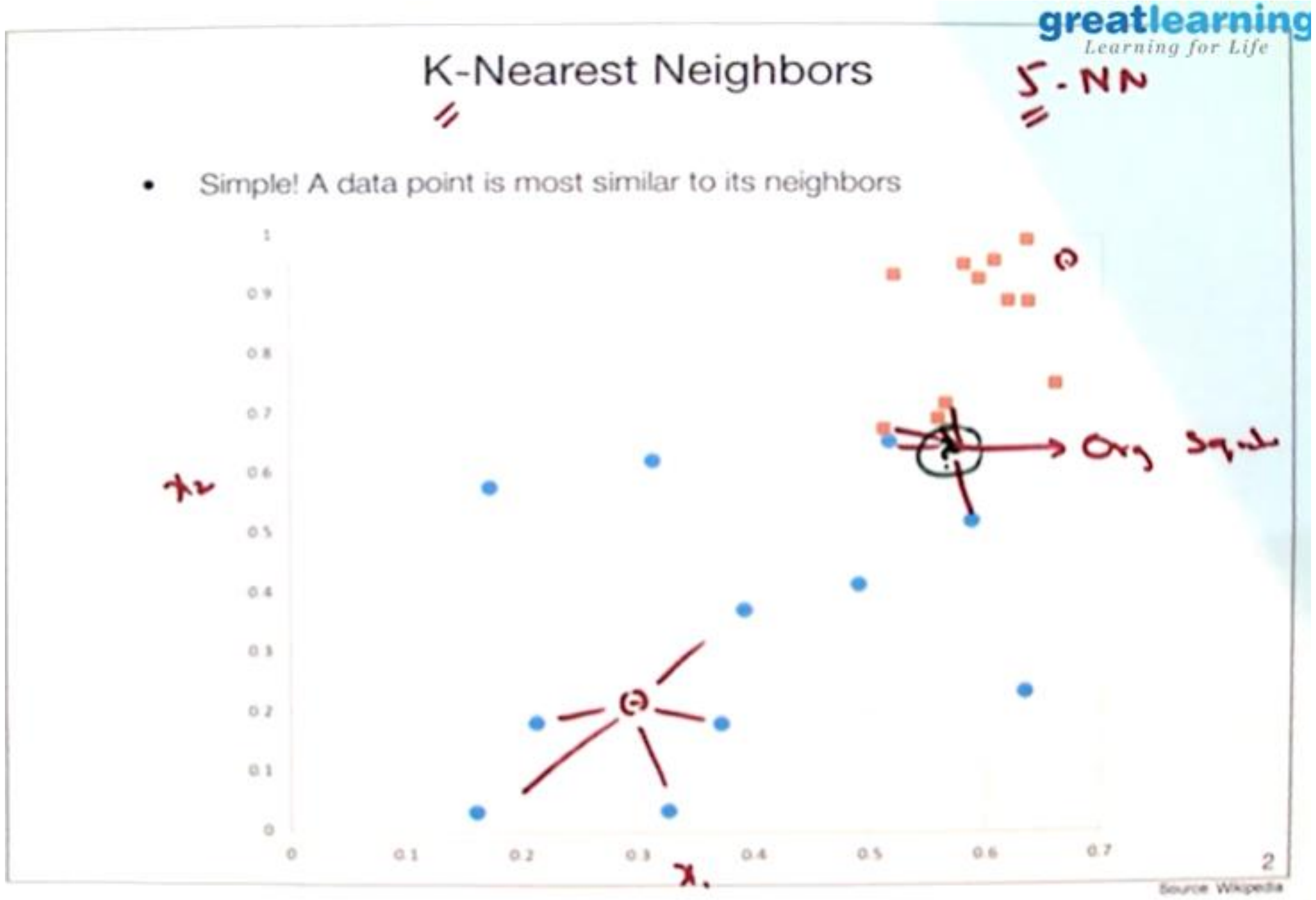
- Advantages
  - Simple, Fast in processing and effective
  - Does well with noisy data and missing data
  - Requires few examples for training (assuming the data set is a true representative of the population)
  - Easy to obtain estimated probability for a prediction
- Dis-advantages
  - Relies on and often incorrect assumption of independent features
  - Not ideal for data sets with large number of numerical attributes
  - Estimated probabilities are less reliable in practice than predicted classes
  - If rare events are not captured in the training set but appears in the test set the probability calculation will be incorrect

Gaussian Naive Bayes classifier

- When some of our independent variables are continuous we cannot calculate conditional probabilities!
- In Gaussian Naive Bayes, continuous values associated with each feature (or independent variable) are assumed to be distributed according to a Gaussian distribution
- All we would have to do is estimate the mean and standard deviation of the continuous variable.



KNN


great learning
Learning for Life
K-Nearest Neighbors
5-NN
Simple! A data point is most similar to its neighbors
Org Signal
x2
x1
2
Source Wikipedia

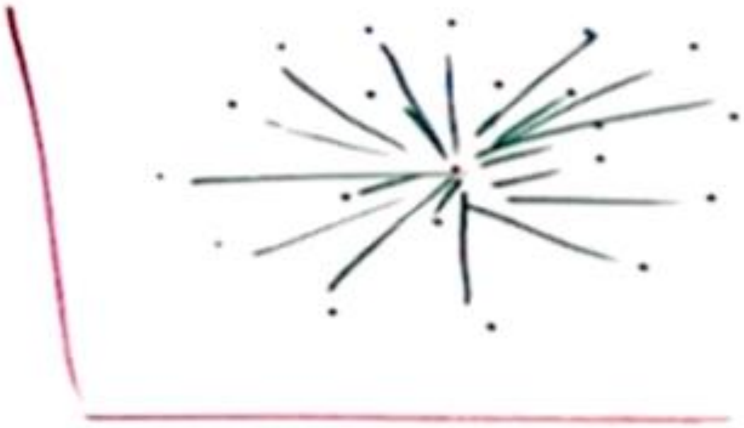
- → define dist
- → define K
- → complexity

euclidean dist

$$\sqrt{(x_1-x_2)^2+(y_1-y_2)^2}$$

manhattan dist

$$|x_2-x_1|+|y_2-y_1|$$



greatlearning
Learning for Life

Distance measure is important

- Most commonly distance is measured using Euclidian distances
- We should always Normalize data
- Other distance measurement methods include
    - Manhattan distance
    - Minkowski distance
    - Mahalanobis distance
    - Cosine similarity

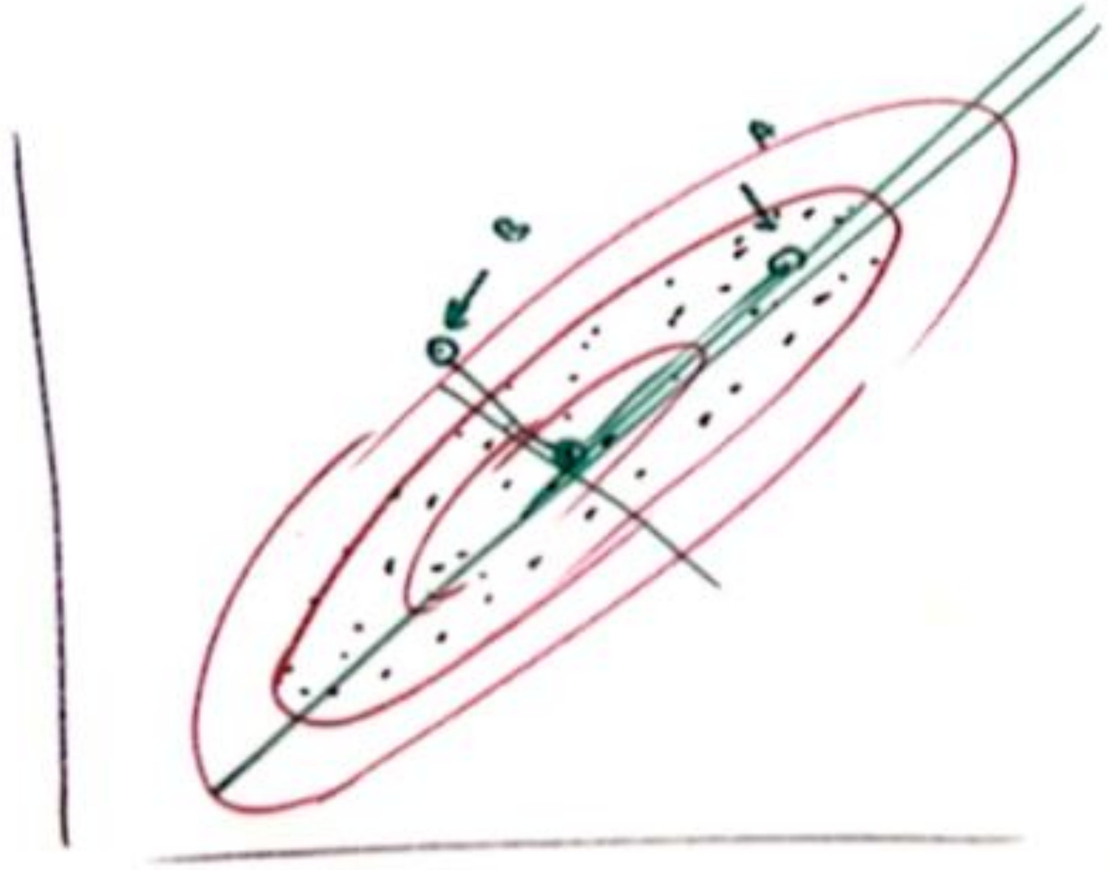
b

a

$$\sum |a_i - b_i|^p$$

$$\sum (a_i - b_i)^2 \rightarrow \text{Euc } (p=2)$$

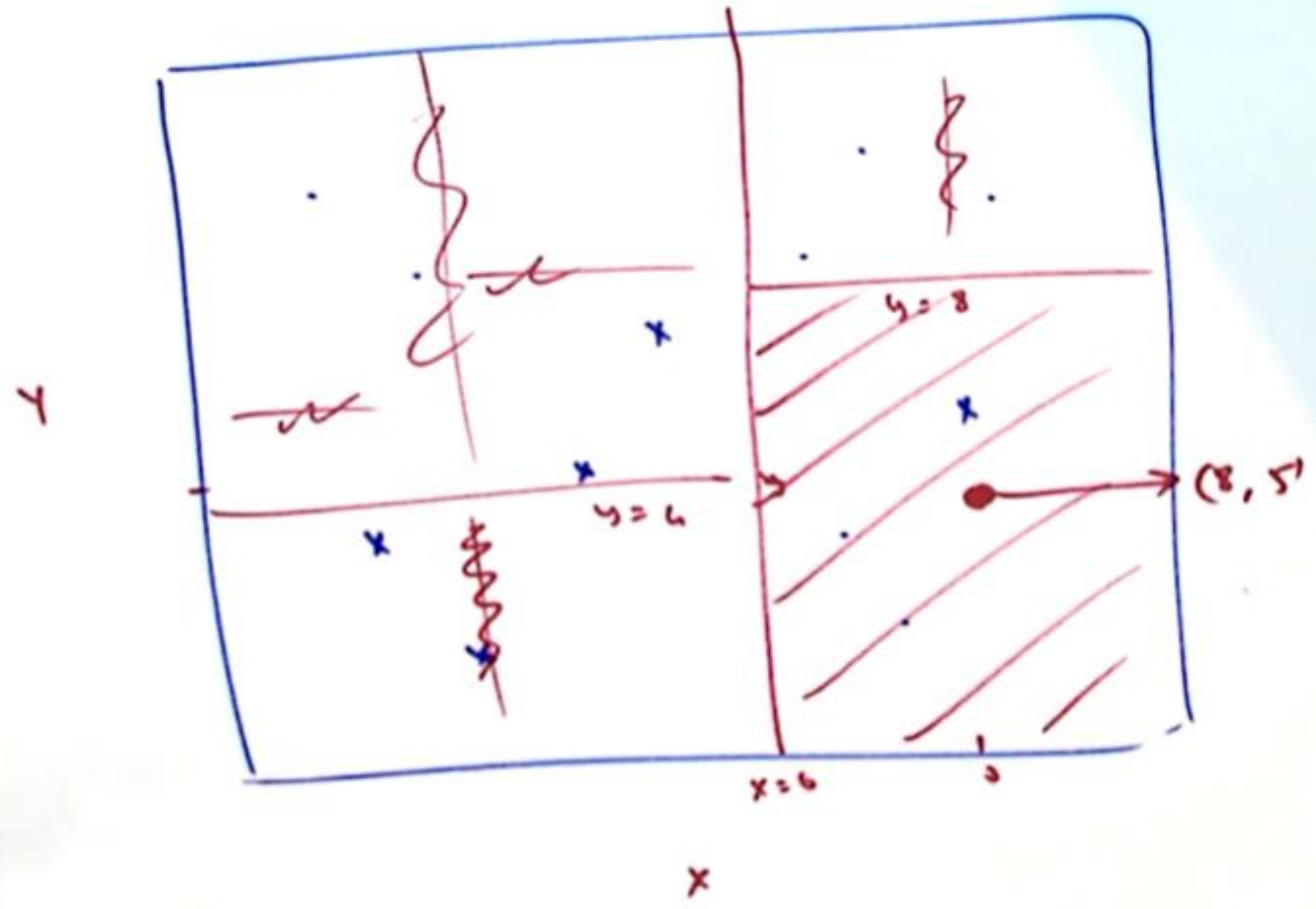
$$\sum |a_i - b_i| \rightarrow \text{man } (p=1)$$


A
B

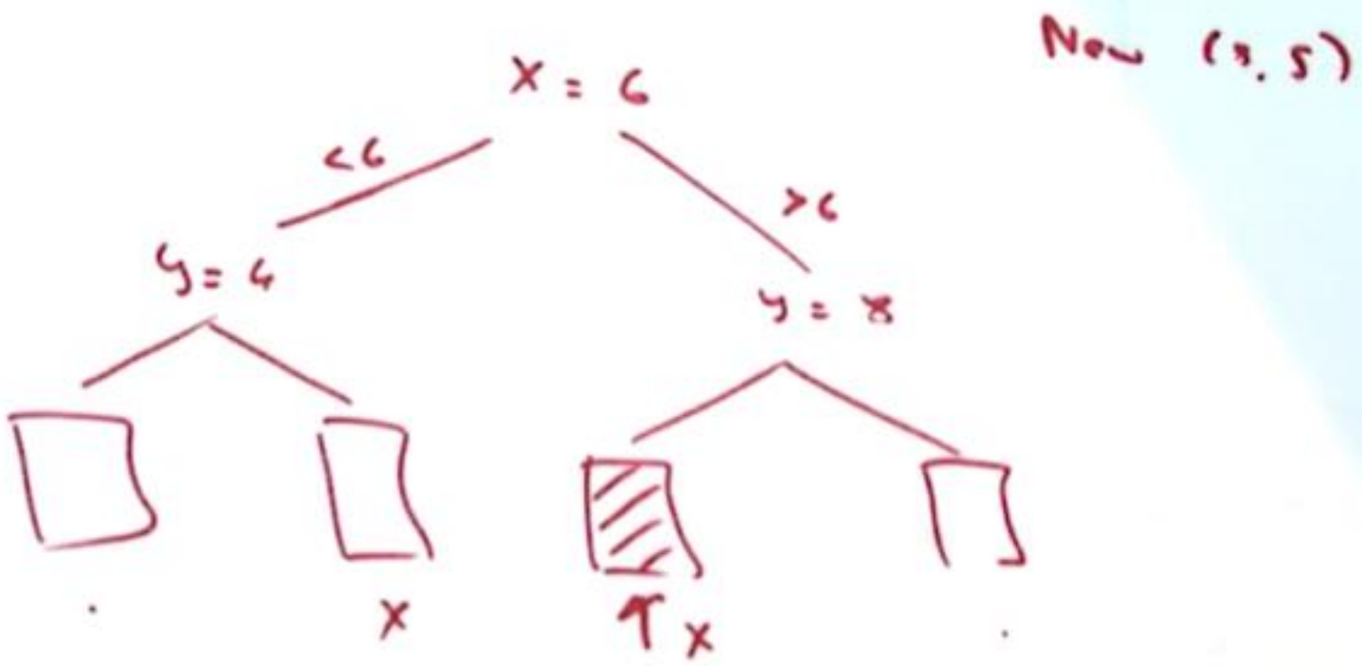
Other Variants

- Radius Neighbor Classifier
  - implements learning based on number of neighbors within a fixed radius r of each training point, where r is a floating point value specified by the user
  - may be a better choice when the sampling is not uniform. However, when there are many attributes and data is sparse, this method becomes ineffective due to curse of dimensionality

- KD Tree nearest neighbor
  - Approach helps reduce the computation time.
  - Very effective when we have large data points but still not too many dimensions



greatlearning
Learning for Life
y = 8
y = 6
(8, 5)
x = 6
Y
X


New (3, 5)
X = 6
<6
>6
y = 4
y = 8
x
x

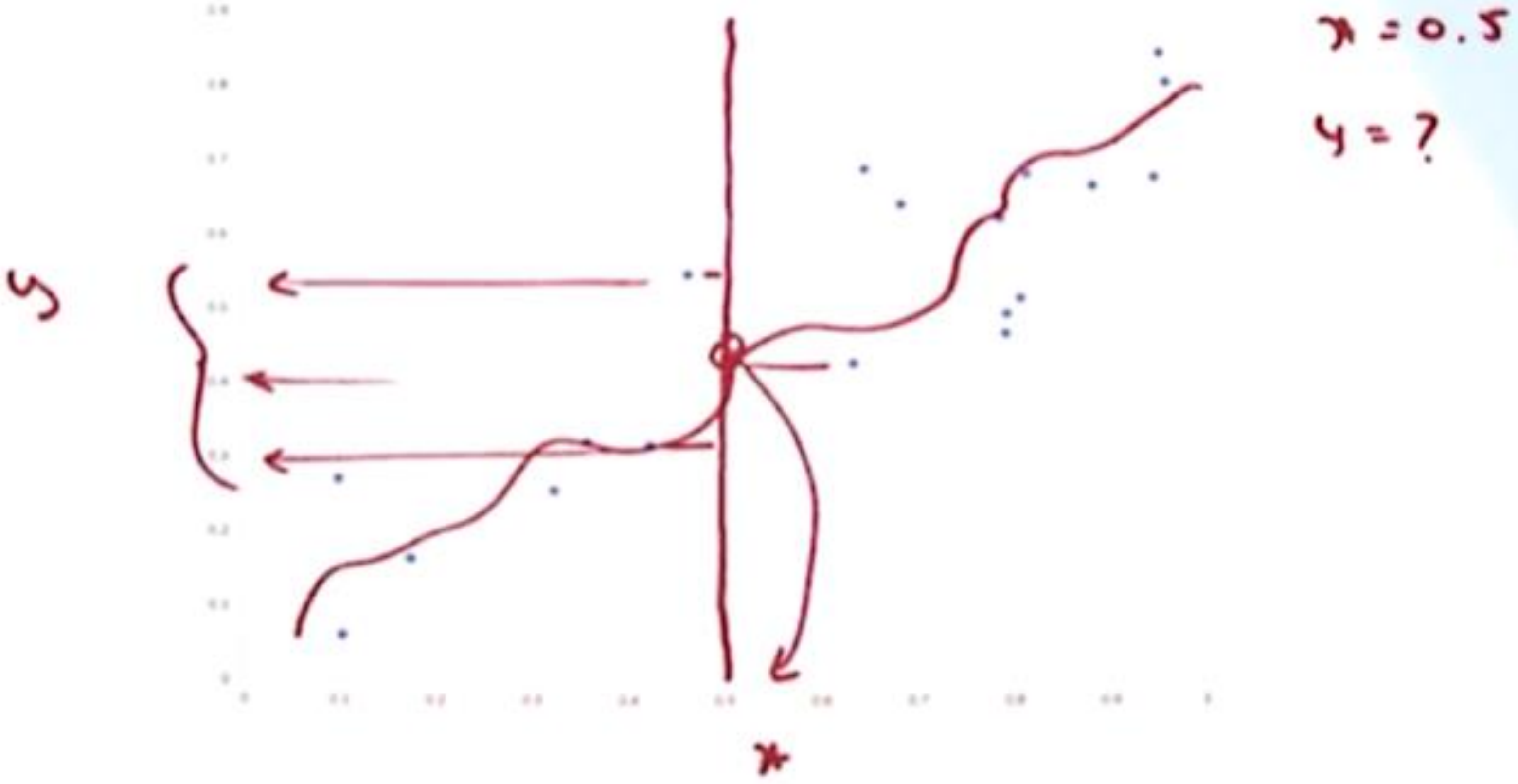
K-NN

- It does not construct a "model". Known as a non-parametric method.
- Classification is computed from a simple majority vote of the nearest neighbors of each point
- Suited for classification where relationship between features and target classes is numerous, complex and difficult to understand and yet items in a class tend to be fairly homogenous on the values of attributes
- Not suitable if the data is too noisy and the target classes do not have clear demarcation in terms of attribute values
- Can also be used for regression

K-NN for regression

3 - NN

- The Neighbors based algorithm can also be used for regression where the labels are continuous data and the label of query point can be average of the labels of the neighbors



K Nearest Neighbors - pros and cons

- Advantages
  - Makes no assumptions about distributions of classes in feature space
  - Can work for multi classes simultaneously
  - Easy to implement and understand
  - Not impacted by outliers
- Dis-advantages
  - Fixing the optimal value of K is a challenge
  - Will not be effective when the class distributions overlap
  - Does not output any models. Calculates distances for every new point (lazy learner)
  - Computationally intensive