
CSE 151A - Discussion 03

Quick Review

Conditional Independence

Events A and B are conditionally independent given C if:

$$P(A, B|C) = P(A|C) \times P(B|C)$$

$$P(A|B, C) = P(A|C)$$

$$P(B|A, C) = P(B|C)$$

Mutual Conditional Independence

Events $X_1 \dots X_d$ are mutually conditionally independent given Y if:

$$P(X_1 \dots X_d|Y) = P(X_1|Y) \times \dots \times P(X_d|Y)$$

Naïve Bayes Classifier : Predict class y_i that maximizes $P(Y = y_i|\vec{X} = \vec{x})$ for some input $\vec{x} = x_1 \dots x_d$.

Like before, $\operatorname{argmax}_{y_i} P(Y = y_i|\vec{X} = \vec{x})$ is equivalent to $\operatorname{argmax}_{y_i} P(\vec{X} = \vec{x}|Y = y_i)P(Y = y_i)$

Assuming mutual conditional independence of $\vec{X} = X_1 \dots X_d$ given Y :

$$\operatorname{argmax}_{y_i} P(Y = y_i|\vec{X} = \vec{x}) = \operatorname{argmax}_{y_i} P(X_1 = x_1|Y = y_i) \times \dots \times P(X_d = x_d|Y = y_i) \times P(Y = y_i)$$

Logarithm Properties

We can use log probabilities to prevent underflow issues by computing sums instead of products:

$$\log [P(X_1 = x_1|Y = y_i) \times \dots \times P(X_d = x_d|Y = y_i) \times P(Y = y_i)]$$

$$= \log \left[\prod_{j=1}^d P(X_j = x_j|Y = y_i) \times P(Y_i = y_i) \right]$$

$$= \left[\sum_{j=1}^d \log P(X_j = x_j|Y = y_i) \right] + \log P(Y_i = y_i)$$

Linear Regression

With $H(x) = w_1x + w_0$,

$$MSE = R_{sq}(w_1, w_0) = \frac{1}{n} \sum_{i=1}^n ((w_1x + w_0) - y_i)^2$$

Least squares solution,

$$w_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad w_0 = \bar{y} - w_1\bar{x}$$

Linear Algebra

Problem 1.

Given the following events, use intuition to determine if A and B are conditionally independent given C :

- a) A : Are people using umbrellas?
 B : Is the sidewalk wet?
 C : Is it raining outside?
- b) A : Was there an earthquake?
 B : Was there a burglary?
 C : Did the house alarm go off?

Problem 2.

A local hospital has collected the symptoms and diagnoses of 20 sick patients. Your friend is sick with a fever, dry cough, and a sore throat, but they are not sneezing. Use the data below and a Naïve Bayes classifier to predict whether your friend has COVID-19, the Cold, or the Flu. Be sure to clearly state your prediction as one of the three illnesses. Show your work.

Patient #	Illness	Fever	Dry Cough	Sore Throat	Sneezing
1	COVID-19	Yes	Yes	No	No
2	COVID-19	Yes	Yes	Yes	No
3	COVID-19	Yes	Yes	No	No
4	Cold	No	Yes	Yes	Yes
5	Cold	No	Yes	Yes	No
6	Cold	No	No	Yes	Yes
7	Cold	No	No	No	Yes
8	Cold	Yes	Yes	Yes	Yes
9	Cold	No	Yes	Yes	Yes
10	Cold	No	No	Yes	Yes
11	Cold	No	No	Yes	Yes
12	Cold	No	Yes	Yes	Yes
13	Cold	No	Yes	No	Yes
14	Flu	Yes	No	No	No
15	Flu	Yes	Yes	Yes	No
16	Flu	Yes	Yes	Yes	No
17	Flu	Yes	Yes	Yes	No
18	Flu	No	No	Yes	No
19	Flu	Yes	Yes	Yes	No
20	Flu	Yes	Yes	No	No

Problem 3.

Suppose that 1% of all computer science students cheat and that a particular cheat-detecting software has a 5% false positive rate and a 10% false negative rate.

Let $A \in \{0, 1\}$ indicate whether a student is cheating, and let $B \in \{0, 1\}$ indicate the result of the cheat-detecting software.

- a) Using this information, determine the probability tables for $P(A)$ and $P(B|A)$.
 $P(A = 1) =$ $P(A = 0) =$

A	$P(B = 1 A)$	$P(B = 0 A)$
0		
1		

- b) The cheat-detecting software just reported that a student named Luke tested positive for cheating. What is the probability that Luke actually cheated?

Problem 4.

x	y
4	1
5	6
9	8

Here is a dataset of (x, y) pairs:

- a) Given the above data, fit a linear prediction rule by computing w_0 and w_1 .
- b) Use the prediction rule from part a) to estimate the y value that corresponds to an input of $x = 2$.
- c) Calculate the MSE of the prediction rule from part a) with respect to the given (x, y) pairs.