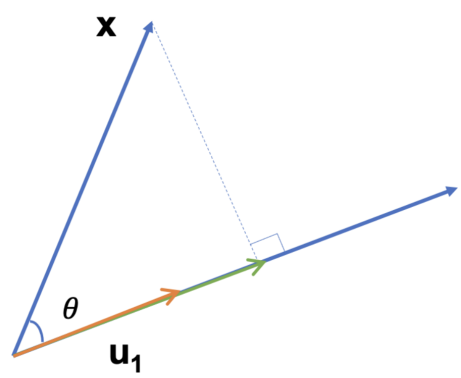
Problem 1

Problem Statement

As the figure shown below, there are two vectors $\boldsymbol{x}$ and $\boldsymbol{u_1}$ (orange color), where $\boldsymbol{u_1}$ has unit length. The projection of $\boldsymbol{x}$ onto $\boldsymbol{u_1}$ is represented by the vector in green. Show that the length of the projection of $\boldsymbol{x}$ onto $\boldsymbol{u_1}$ has length $\boldsymbol{u_1}^T\boldsymbol{x}$



My solution:
Fact 1: The vector projection is the unit vector $\frac{\boldsymbol{u_1}}{|\boldsymbol{u_1}|}$ times the scalar projection of $\boldsymbol{x}$ onto $\boldsymbol{u_1}$,

$$Proj_{\boldsymbol{u_1}}\boldsymbol{x} = \frac{\boldsymbol{x} \cdot \boldsymbol{u_1}}{|\boldsymbol{u_1}|^2} \cdot \boldsymbol{u_1}$$

NOTE: $|\boldsymbol{u_1}|^2 = 1$ since u_1 is the unit vector.
Fact 2: We need to establish that $a \cdot b = a^T \cdot b$
To see this to be true, note that $\boldsymbol{a} \cdot \boldsymbol{b} = a_1b_1 + a_2b_2 + \ldots + a_nb_n$ for $\boldsymbol{a}$ is the $n \times 1$ column vector, thus $\boldsymbol{a}^T$ will be the $1 \times n$ row vector. Then $\boldsymbol{b}$ is the $n \times 1$ column vector, thus

$$\boldsymbol{a}^T \cdot \boldsymbol{b} = [a_1b_1 + \ldots + a_nb_n]$$

is the 1×1 matrix, a scalar. Therefore, $\boldsymbol{a} \cdot \boldsymbol{b} = \boldsymbol{a}^T\boldsymbol{b}$.
Using the facts established above,

$$Proj_{\boldsymbol{u_1}}\boldsymbol{x} = \frac{\boldsymbol{x} \cdot \boldsymbol{u_1}}{|\boldsymbol{u_1}|^2} \cdot \boldsymbol{u_1}$$

However, $|\boldsymbol{u_1}|^2 = 1$ thus,

$$Proj_{\boldsymbol{u_1}}\boldsymbol{x} = (\boldsymbol{x} \cdot \boldsymbol{u_1}) \cdot \boldsymbol{u_1}$$

Then using the fact that, $\boldsymbol{a} \cdot \boldsymbol{b} = \boldsymbol{a}^T\boldsymbol{b}$ and because the dot product is commutative, we have

$$Proj_{\boldsymbol{u_1}}\boldsymbol{x} = (\boldsymbol{u_1}^T\boldsymbol{x}) \cdot \boldsymbol{u_1}$$

The length of this projection vector is the scalar $|\boldsymbol{u_1}^T\boldsymbol{x}|$, however, given that $\boldsymbol{u_1}$ is the unit vector that is in the same direction as the projection vector, our scalar projection will be positive. We then have that the length of the projection $\boldsymbol{x}$ onto $\boldsymbol{u_1}$ is $\boldsymbol{u_1}^T\boldsymbol{x}$, as desired.

Problem 2

Problem Statement

Show that,

$$\frac{1}{N}\sum_{i=1}^{N}(\boldsymbol{u_1}^T\boldsymbol{x_i} - \boldsymbol{u_1}^T\bar{\boldsymbol{x}})^2 = \boldsymbol{u_1}^T\boldsymbol{S}\boldsymbol{u_1}$$

Such that $\boldsymbol{S}$ is the data covariance matrix defined by,

$$\boldsymbol{S} = \frac{1}{N}\sum_{i=1}^{N}(\boldsymbol{x_i} - \bar{\boldsymbol{x}})(\boldsymbol{x_i} - \bar{\boldsymbol{x}})^T$$

My Solution:

First: Let us expand

$$\Rightarrow \frac{1}{N}\sum_{i=1}^{N}(\boldsymbol{u_1}^T\boldsymbol{x_i} - \boldsymbol{u_1}^T\bar{\boldsymbol{x}})(\boldsymbol{u_1}^T\boldsymbol{x_i} - \boldsymbol{u_1}^T\bar{\boldsymbol{x}})$$

Second: Let us group terms together to get the form of the inner product that is, given a and b are column vectors then $\boldsymbol{a}\cdot\boldsymbol{b} = \boldsymbol{a}^T\cdot\boldsymbol{b}$ this will allow us to rearrange nicely so that we can deal with scalar multiplication (where order doesn't matter),

$$\Rightarrow \frac{1}{N}\sum_{i=1}^{N}(\boldsymbol{u_1}^T(\boldsymbol{x_i} - \bar{\boldsymbol{x}}))((\boldsymbol{x_i} - \bar{\boldsymbol{x}})^T\boldsymbol{u_1})$$

Third: Deploy the definition of the covariance matrix and the fact that scalar multiplication is commutative,

$$\Rightarrow \boldsymbol{u_1}^T(\frac{1}{N}\sum_{i=1}^{N}(\boldsymbol{x_i} - \bar{\boldsymbol{x}})(\boldsymbol{x_i} - \bar{\boldsymbol{x}})^T)\boldsymbol{u_1}$$

Fourth: Rewrite in matrix notation,

$$\frac{1}{N}\sum_{i=1}^{N}(\boldsymbol{u_1}^T\boldsymbol{x_i} - \boldsymbol{u_1}^T\bar{\boldsymbol{x}})^2 = \boldsymbol{u_1}^T\boldsymbol{S}\boldsymbol{u_1}$$

Which is the desired result.

Problem 3

Problem Statement

For a binary classification problem with three binary features, the data is shown below:

a	b	c	K
1	0	1	1
1	1	1	1
0	1	1	0
1	1	0	0
1	0	1	0
0	0	0	1
0	0	0	1
0	0	1	0

(a) Based on the naive Bayes classifier, what is $P(K = 1|a = 1, b = 1, c = 0)$?

(b) Based on the naive Bayes classifier, what is $P(K = 0|a = 1, b = 1)$?

My Solution:

PART A

In the case of muliple independent evidences, $E_1, E_2, \ldots, E_n$, the Naive Bayes is:

$$P(H|E_1, E_2, \ldots, E_n) = \frac{P(E_1|H)P(E_2|H)\cdots P(E_n|H)P(H)}{\sum_H P(E_1|H)P(E_2|H)\ldots P(E_n|H)P(H)} \quad (2)$$

Usually, we are only interested in calculating the numerator, since the denominator is a normalization constant (once $E_1, E_2, \ldots, E_n$ are given, the denominator is fixed.

$$P(a = 1|K = 1) = \frac{2}{4} \Rightarrow P(b = 1|K = 1) = \frac{1}{4} \Rightarrow P(c = 0|K = 1) = \frac{2}{4} \Rightarrow P(K = 1) = \frac{4}{8}$$

$$P(K = 1|a = 1, b = 1, c = 0) \propto \frac{2}{4} \cdot \frac{1}{4} \cdot \frac{2}{4} \cdot \frac{4}{8} = \frac{1}{32} \approx 0.03125$$

Now we find $P(K = 0|a = 1, b = 1, c = 0)$,

$$P(a = 1|K = 0) = \frac{2}{4} \Rightarrow P(b = 1|K = 0) = \frac{2}{4} \Rightarrow P(c = 0|K = 0) = \frac{1}{4} \Rightarrow P(K = 0) = \frac{4}{8}$$

$$P(K = 0|a = 1, b = 1, c = 0) \propto \frac{2}{4} \cdot \frac{2}{4} \cdot \frac{1}{4} \cdot \frac{4}{8} = \frac{1}{32} \approx 0.03125$$

Then,

$$P(K = 1|a = 1, b = 1, c = 0) = \frac{P(a = 1|K = 1)P(b = 1|k = 1)P(c = 0|K = 1)P(K = 1)}{P(K = 1|a = 1, b = 1, c = 0) + P(K = 0|a = 1, b = 1, c = 0)} = \frac{\frac{1}{32}}{\frac{1}{32} + \frac{1}{32}}$$

$$P(K = 1|a = 1, b = 1, c = 0) = \frac{1}{2}$$

PART B

$$P(a=1|K=0)=\frac{2}{4}\Rightarrow P(b=1|k=0)=\frac{2}{4}\Rightarrow P(K=0)=\frac{4}{8}$$

$$P(K=0|a=1,b=1)\propto\frac{2}{4}\cdot\frac{2}{4}\cdot\frac{4}{8}=\frac{1}{8}=0.125$$

Now we find $P(K=1|a=1,b=1)$,

$$P(a=1|K=1)=\frac{2}{4}\Rightarrow P(b=1|k=1)=\frac{1}{4}\Rightarrow P(K=1)=\frac{4}{8}$$

$$P(K=1|a=1,b=1)\propto\frac{2}{4}\cdot\frac{1}{4}\cdot\frac{4}{8}=\frac{1}{16}=0.0625$$

Then,

$$P(K=0|a=1,b=1)=\frac{P(a=1|k=0)P(b=1|K=0)P(K=0)}{P(K=0|a=1,b=1)+P(K=1|a=1,b=1)}=\frac{\frac{1}{8}}{\frac{1}{8}+\frac{1}{16}}=\frac{2}{3}$$