Machine Learning HW1

Wenhan Yang

Due: February 20 2021

Question 1

(a)

Using the conditional probability formula:

$$\begin{aligned}
&\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}, \mathbb{P}(B|A) = \frac{\mathbb{P}(B \cap A)}{\mathbb{P}(A)} = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)} \\
\therefore &\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B|A) \\
\therefore &\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(A)\mathbb{P}(B|A)}{\mathbb{P}(B)}
\end{aligned}$$

(b)

Notation : $\mathbb{P}(A, B, C) = \mathbb{P}(A \cap B \cap C)$

$$\begin{aligned}
\mathbb{P}(A \cap B \cap C) &= \mathbb{P}((A \cap B) \cap C) \\
&= \mathbb{P}((A \cap B)|C)\mathbb{P}(C) \\
&= \mathbb{P}(A|(B \cap C))\mathbb{P}(B|C)\mathbb{P}(C) \\
&= \mathbb{P}(A|(B, C))\mathbb{P}(B|C)\mathbb{P}(C) \\
&\text{or} \\
\mathbb{P}(A \cap B \cap C) &= \mathbb{P}((A \cap B) \cap C) = \mathbb{P}(C \cap (A \cap B)) \\
&= \mathbb{P}(C|(A \cap B))\mathbb{P}(A \cap B) \\
&= \mathbb{P}(C|(A \cap B))\mathbb{P}(B|A)\mathbb{P}(A) \\
&= \mathbb{P}(C|(A, B))\mathbb{P}(B|A)\mathbb{P}(A)
\end{aligned}$$

(c)

$\mathbb{E}[X] = 1 \cdot \mathbb{P}(\text{A occurs})) + 0 \cdot \mathbb{P}(\text{A not occur})) = 1 \cdot \mathbb{P}(A) + 0 \cdot \mathbb{P}(A^C) = \mathbb{P}(A)$
$\therefore \mathbb{E}[X] = \mathbb{P}(A)$

(d)

(i)

$$\begin{aligned}
\mathbb{P}(X=0,Y=0) &= \mathbb{P}(X=0,Y=0,Z=0)+\mathbb{P}(X=0,Y=0,Z=1)\\
&= \frac{1}{15}+\frac{4}{15}=\frac{1}{3}\\
\mathbb{P}(X=0) &= \mathbb{P}(X=0,Y=0,Z=0)+\mathbb{P}(X=0,Y=0,Z=1)\\
&\quad +\mathbb{P}(X=0,Y=1,Z=0)+\mathbb{P}(X=0,Y=1,Z=1)\\
&= \frac{1}{15}+\frac{4}{15}+\frac{1}{10}+\frac{8}{45}=\frac{11}{18}\\
\mathbb{P}(Y=0) &= \mathbb{P}(X=0,Y=0,Z=0)+\mathbb{P}(X=0,Y=0,Z=1)\\
&\quad +\mathbb{P}(X=1,Y=0,Z=0)+\mathbb{P}(X=1,Y=0,Z=1)\\
&= \frac{1}{15}+\frac{4}{15}+\frac{1}{15}+\frac{2}{15}=\frac{8}{15}\\
\mathbb{P}(X=0)\mathbb{P}(Y=0) &= \frac{11}{18}\cdot\frac{8}{15}=\frac{44}{135}\neq\frac{1}{3}=\mathbb{P}(X=0,Y=0)
\end{aligned}$$

$\therefore$ X is not independent of Y.

(ii)

$$\begin{aligned}
\mathbb{P}(X=0,Y=0|Z=0) &= \frac{1}{15}\\
\mathbb{P}(X=0|Z=0) &= \mathbb{P}(X=0,Y=0|Z=0)+\mathbb{P}(X=0,Y=1|Z=0)\\
&= \frac{1}{15}+\frac{1}{10}=\frac{1}{6}\\
\mathbb{P}(Y=0|Z=0) &= \mathbb{P}(X=0,Y=0|Z=0)+\mathbb{P}(X=1,Y=0|Z=0)\\
&= \frac{1}{15}+\frac{1}{15}=\frac{2}{15}\\
\mathbb{P}(X=0|Z=0)\mathbb{P}(Y=0|Z=0) &= \frac{1}{6}\cdot\frac{2}{15}=\frac{1}{45}\neq\frac{1}{15}=\mathbb{P}(X=0,Y=0|Z=0)
\end{aligned}$$

$\therefore$ X is not conditionally independent of Y given Z.

(iii)

$$\begin{aligned}
\mathbb{P}(X=0|X+Y>0) &= \frac{\mathbb{P}(X=0,X+Y>0)}{\mathbb{P}(X+Y>0)}=\frac{\mathbb{P}(X=0,Y>0)}{\mathbb{P}(X+Y>0)}\\
&= \frac{\mathbb{P}(X=0,Y=1)}{\mathbb{P}(X=0,Y=1)+\mathbb{P}(X=1,Y=0)+\mathbb{P}(X=1,Y=1)}\\
&= \frac{\frac{1}{10}+\frac{8}{45}}{\frac{1}{10}+\frac{8}{45}+\frac{1}{15}+\frac{2}{15}+\frac{1}{10}+\frac{4}{45}}=\frac{5}{12}
\end{aligned}$$

Question 2

*Go to the directory where you save problem-2.py using cd path_name
*To run problem-2.py, use command python problem-2.py run in your terminal.

(a)

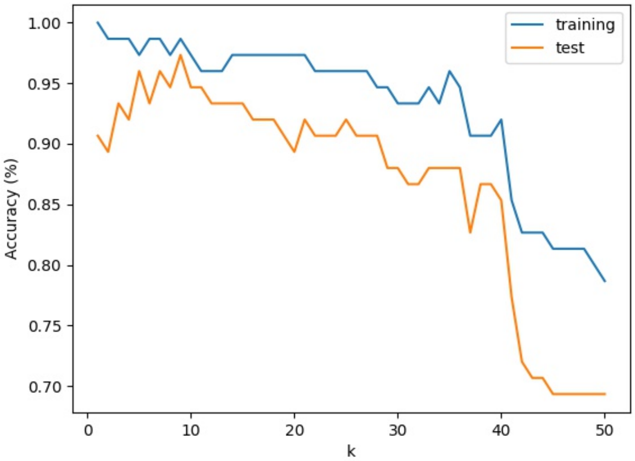
Class 0 has 50 elements.
Class 1 has 50 elements.
Class 2 has 50 elements.

(b)

Accuracy = 100%
This accuracy is **not** meaningful. This accuracy is of the training set, where the model could over-fit to the training data, i.e. remembering the class of every data. In this case, k=1, the 1'st nearest neighbour of the point is actually the point itself, which is not meaningful.

(c)

The optimal k value is: 9



**This graph can be found in the same folder where you save problem-2.py, and named "wy818_accuracy.jpg"*

(d)

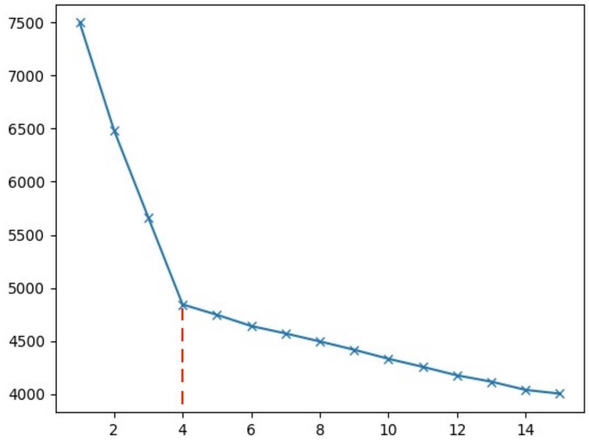
Predicted class of this plant: 0

Question 3

*Go to the directory where you save problem-3.py using cd path_name
*To run problem-3.py, use command python problem-3.py run in your terminal.

(a)

The elbow point occurs at k=4. Hence, 4 clusters should be used for this data.



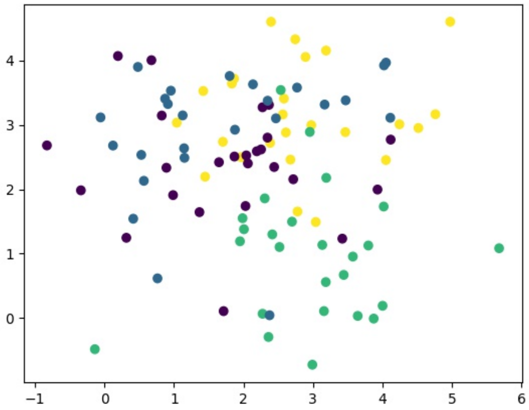
**This graph can be found in the same folder where you save problem-3.py, and named ”wy818_elbowcurve.jpg”*

(b)

Cluster 1 has 25 observations.
Cluster 2 has 25 observations.
Cluster 3 has 25 observations.
Cluster 4 has 25 observations.
Value of inertia is: 4844.925818

(c)

From the graph, it is **not** a good clustering. We are only using the first 2 variables, hence the scatter plot might not be reliable.



**This graph can be found in the same folder where you save problem-3.py, and named ”wy818_scatterplot.jpg”*