HOMEWORK 1: MACHINE LEARNING REVIEW

Data Science II

Devashi Gulati
University of Georgia

Contents

1 Conditional Probability and the Chain Rule **3**

1.1 3

1.2 3

2 Total Probability **3**

2.1 3

2.2 4

3 Naive Bayes **5**

3.1 5

3.2 5

4 Logistic Regression **6**

4.1 6

4.2 6

5 Coding **6**

1 Conditional Probability and the Chain Rule

1.1

Given Conditional Probability definition:

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

To Prove:

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

From the definition of Conditional Probability, $P(A \cap B) = P(A|B)P(B)$.

Using this, we get $P(A \cap B \cap C) = P(A|B, C)P(B \cap C)$. Here $P(A|B, C)$ is the probability of event A occurring, assuming event B and event C occurred. Hence $P(A \cap B \cap C) = P(A|B, C)P(B \cap C) = P(A|B, C)P(B|C)P(C)$.

1.2

Bayes' Theorem :

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

From the definition of Conditional Probability, $P(A \cap B) = P(A|B)P(B) = P(B|A)P(A)$. Hence $P(A|B) = \frac{P(B|A)P(A)}{P(B)}$

- $P(A|B)$: Probability of B occurring, assuming A has occurred.
- $P(B|A)$: Probability of A occurring, assuming B has occurred.
- $P(A)$: Probability of A occurring
- $P(B)$: Probability of B occurring

2 Total Probability

1. Coin Flip such that $P(Coin = Heads) = p$

   2a If Coin = Heads then Fair Die roll, so $P(dice = x) = 1/6$

   2b If Coin = Tails then Loaded Die roll such that

   - $P(dice = 6) = 1/2$
   - if $x \in \{1, 2, 3, 4, 5\}$ then $P(dice = x) = 1/10$

2.1

Expected die roll in terms of p:

$$\begin{aligned}
E[X] &= \sum_i x_i P(X = x_i) \\
&= p \sum_{j=1}^{6} j \frac{1}{6} + (1-p)\left(\sum_{k=1}^{5} k \frac{1}{10} + 6\frac{1}{2} \right) \\
&= p\frac{1}{6}21 + (1-p)\left(15\frac{1}{10} + \frac{6}{2}\right) \\
&= p\frac{7}{2} + (1-p)\left(\frac{3}{2} + \frac{6}{2}\right) \\
&= \frac{7p}{2} + \frac{(1-p)9}{2} \\
&= \frac{7p + 9 - 9p}{2} \\
&= \frac{9 - 2p}{2} \\
&= \frac{9}{2} - p
\end{aligned}$$

2.2

Variance of die roll in terms of p:

$$Var(X) = E[X^2] - (E[X])^2$$

$$\begin{aligned}
E[X^2] &= \sum_i (x_i)^2 P(X = x_i) \\
&= p \sum_{j=1}^{6} j^2 \frac{1}{6} + (1-p)\left(\sum_{k=1}^{5} k^2 \frac{1}{10} + 6^2\frac{1}{2} \right) \\
&= p\frac{1}{6}(1+4+9+16+25+36) + (1-p)\left((1+4+9+16+25)\frac{1}{10} + \frac{36}{2}\right) \\
&= p\frac{1}{6}91 + (1-p)\left(55\frac{1}{10} + \frac{36}{2}\right) \\
&= \frac{91p}{6} + (1-p)\left(\frac{11}{2} + \frac{36}{2}\right) \\
&= \frac{91p}{6} + (1-p)\frac{47}{2} \\
&= \frac{91p}{6} + \frac{47 - 47p}{2} \\
&= \frac{91p}{6} + \frac{141 - 141p}{6} \\
&= \frac{141 - 50p}{6}
\end{aligned}$$

$$\begin{aligned}
Var(X) &= E[X^2] - (E[X])^2 \\
&= \frac{141 - 50p}{6} - \left(\frac{9 - 2p}{2}\right)^2 \\
&= \frac{141 - 50p}{6} - \frac{81 - 36p + 4p^2}{4} \\
&= \frac{282 - 100p}{12} - \frac{243 - 108p + 12p^2}{12} \\
&= \frac{282 - 243 + 108p - 100p - 12p^2}{12} \\
&= \frac{39 + 8p - 12p^2}{12}
\end{aligned}$$

3 Naive Bayes

$$f : X = (x_1, x_2, \dots, x_n) \longrightarrow Y = \{T, F\}$$

- x_1 : boolean attribute
- $x_i, i \in \{2, \dots, n\}$: continuous attribute

3.1

$x_i, i \in \{2, \dots, n\}$: continuous attributes are Gaussian

Give and briefly explain the total number of parameters that you would need to estimate in order to classify a future observation using a Naive Bayes (NB) classifier.

$$\theta_{ij} = P(X = x_i | Y = y_j)$$

A Gaussian distribution needs 2 parameters:

- μ : mean/ expected value
- σ^2 : variance

Assuming Naive Bayes (NB) Classifier, we need 2n parameters, 2 for x_1 and 2 each for $x_i, i \in \{2, \dots, n\}$.

3.2

How many more parameters would be required without the conditional independence assumption? No need for an exact number; an order of magnitude estimate will suffice.

Without the conditional independence assumption, we would need exponential order of parameters i.e $O(2^n)$.

4 Logistic Regression

4.1

Prove the decision boundary for Logistic Regression is linear. i.e., show that $P(Y|X)$ has the form:

$$w_0 + \sum_i w_i X_i$$

Decision Boundary := A decision boundary is the region of a problem space in which the output label of a classifier is ambiguous. It partitions the underlying space into the possible outcomes.

We know that

$$P(Y = 0|X) = \frac{1}{1 + exp(w_0 + \sum_i w_i X_i)}$$

$$\begin{aligned} P(Y = 1|X) &= 1 - P(Y = 0|X) \\ &= 1 - \frac{1}{1 + exp(w_0 + \sum_i w_i X_i)} \\ &= \frac{exp(w_0 + \sum_i w_i X_i)}{1 + exp(w_0 + \sum_i w_i X_i)} \end{aligned}$$

So it is $exp(w_0 + \sum_i w_i X_i)$ in the numerator which determines whether the Logistic Regression predicts $Y = 0$ or $Y = 1$. This is because the decision boundary is

$$P(Y = 0|X) = 0.5 = P(Y = 1|X)$$

This gives the following equation:

$$\begin{aligned} \frac{exp(w_0 + \sum_i w_i X_i)}{1 + exp(w_0 + \sum_i w_i X_i)} = \frac{1}{2} = \frac{1}{1+1} \\ exp(w_0 + \sum_i w_i X_i) = 1 \\ w_0 + \sum_i w_i X_i = 0 \end{aligned}$$

Hence the decision boundary is linear.

4.2

Briefly describe one advantage and one disadvantage of LR compared to NB.

5 Coding

References