Machine Learning Worksheet Solution 2

Julius Jankowski

1 Basic Probability

Problem 1:

The problem describes a scenario which can be described by two random variables $X = T, C$ and $Y = ST, SC$, where X states whether a person is a terrorist (T) or a (upstanding) citizen (C). $p(X)$ and the conditional probabilty $p(Y \mid X)$ are given by the problem description.

The desired solution for the stated problem is the conditional probability $p(X = T \mid Y = ST)$. With Bayes theorem:

$$p(X = T \mid Y = ST) = p(Y = ST \mid X = T) \cdot \frac{p(X = T)}{p(Y = ST)} \tag{1.1}$$

Where $p(Y = ST)$ is unknown in the equation. It can be computed by marginalizing the joint probability distribution $p(X, Y)$. Expressing the joint probabilities by the known quantities results in:

$$p(Y = ST) = \sum_X p(X, Y = ST) = p(Y = ST \mid X = T)p(X = T) + p(Y = ST \mid X = C)p(X = C) \tag{1.2}$$

Fusing equation 1.1 and 1.2 yields $p(X = T \mid Y = ST) \approx 0.161$.

Problem 2:

The solution for problem 2 is the conditional probability $p(Y = 2 \mid X = 3)$, where X relates to the number of red balls randomly drawn from the bowl and Y the number of red balls in the bowl as a result from two coin tosses.

By intuiton, one can build up a joint probability table from the fair coin toss and the probability distribution of red balls according to the number of red balls in the bowl:

		Drawn red balls			
		0	1	2	3
	0	$\frac{1}{4}$	0	0	0
Heads	1	$\frac{1}{16}$	$\frac{3}{16}$	$\frac{3}{16}$	$\frac{1}{16}$
	2	0	0	0	$\frac{1}{4}$

Based on the joint probabilities the solution can be computed:

$$p(Y = 2 \mid X = 3) = \frac{p(X = 3, Y = 2)}{p(X = 3, Y = 0) + p(X = 3, Y = 1) + p(X = 3, Y = 2)} = 0.8 \tag{1.3}$$

following the same ideas as for problem 1.

Problem 3:

At each new decision, there is a probability of 0.5 that the next coin flip results in heads or tails. By iteratively multiplying this probabilites according to the number of occured tails in a row, we get the probility distribution with respect to the number of tails in one run:

$$p(T = t) = (\frac{1}{2})^{t+1} \tag{1.4}$$

It is to the power of $t + 1$ since there is an additional flip required where heads have to show up. Based on this distribution, we can compute the mean:

$$E[T] = \sum_{t=0}^{\infty} t \cdot p(T = t) = \sum_{t=0}^{\infty} t \cdot (\frac{1}{2})^{t+1} \tag{1.5}$$

After changing the sum to the form where the properties of the geometric series is applicable, we get:

$$E[T] = \frac{1}{2} \sum_{t=0}^{\infty} t \cdot (\frac{1}{2})^{t} = 1 \tag{1.6}$$

The expected number of heads is of course $E[H] = 1$.

Problem 4:

Computing the mean:

$$E[X] = \int_{-\infty}^{\infty} x \cdot p(x) dx = \frac{1}{b - a} \int_{a}^{b} x dx = \frac{1}{2}(b + a) \tag{1.7}$$

Computing the variance:

$$\begin{aligned} Var[X] = E[X^2] - E[X]^2 &= \frac{1}{b - a} \int_{a}^{b} x^2 dx - \frac{1}{4}(b + a)^2 = \frac{1}{3} \frac{b^3 - a^3}{b - a} - \frac{1}{4}(b + a)^2 && (1.8) \\ &= \frac{1}{3}(b^2 + ba + a^2) - \frac{1}{4}(b^2 + 2ba + a^2) = \frac{1}{12}(b - a)^2 && (1.9) \end{aligned}$$

Problem 5:

First tower property:

$$E[X] = E_Y[E_{X|Y}[X]] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x \cdot p(x,y)dxdy \tag{1.10}$$

Incorporating Bayes theorem with $p(x,y) = p(x \mid y) \cdot p(y)$:

$$\begin{aligned} E[X] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x \cdot p(x \mid y) \cdot p(y)dxdy = \int_{-\infty}^{\infty}[\int_{-\infty}^{\infty} x \cdot p(x \mid y)dx] \cdot p(y)dy \quad (1.11)\\ = \int_{-\infty}^{\infty} E_{X|Y}[X] \cdot p(y)dy = E_Y[E_{X|Y}[X]] \quad (1.12)\end{aligned}$$

Second tower property:

$$Var[X] = E_Y[Var_{X|Y}[X]] + Var_Y[E_{X|Y}[X]] \tag{1.13}$$

Using the formula for the variance $Var[X] = E[X^2] - E[X]^2$:

$$\begin{aligned} Var[X] = E_Y[E_{X|Y}[X^2]] - E_Y[E_{X|Y}[X]^2] + E_Y[E_{X|Y}[X]^2] - E_Y[E_{X|Y}[X]]^2 \quad (1.14)\\ = E_Y[E_{X|Y}[X^2]] - E_Y[E_{X|Y}[X]]^2 \quad (1.15)\\ = \int_{-\infty}^{\infty} p(y) \cdot \int_{-\infty}^{\infty} p(x \mid y) \cdot x^2 dxdy - (\int_{-\infty}^{\infty} p(y) \cdot \int_{-\infty}^{\infty} p(x \mid y) \cdot xdxdy)^2 \quad (1.16)\\ = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^2 \cdot p(x,y)dxdy - (\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x \cdot p(x,y)dxdy)^2 = E[X^2] - E[X]^2 \quad (1.17)\end{aligned}$$

2 Probability Inequalities

Problem 6:

Proving the formula $p(|\frac{1}{n}\sum_{i=1}^{n} X_i - E[X_i]| > \epsilon) \to 0$ as $n \to \infty$.

Since all random variables are i.i.d., the mean is equal for all variables:

$$E[X_i] = E[X_j] \forall i, j \tag{2.1}$$

This implies that the expectation for an average X over n variables is equal as well:

$$E[X] = E[\frac{1}{n}\sum_{i=1}^{n} X_i] = E[X_i] \forall i \tag{2.2}$$

With this findings we can reformulate the original formula:

$$p(|(\frac{1}{n}\sum_{i=1}^{n} X_i) - E[X]| > \epsilon) = p(|X - E[X]| > \epsilon) \leq \frac{Var[X]}{\epsilon^2} \tag{2.3}$$

according to Chebyshev's inequality, where $X = \frac{1}{n}\sum_{i=1}^{n} X_i$. Now in order to prove the original formula, we have to prove that $Var[X]$ tends to zero as n tends to infinity. Since the variance of all X_i should be equal due to i.i.d., we incorporate a new variable λ for this quantity:

$$\lambda = Var[X_i] \forall i \tag{2.4}$$

Now we can solve for the variance of X:

$$Var[X] = Var[\frac{1}{n}\sum_{i=1}^{n} X_i] = \frac{1}{n^2} \cdot Var[\sum_{i=1}^{n} X_i] = \frac{1}{n^2} \cdot \sum_{i=1}^{n} Var[X_i] = \frac{\lambda}{n} \tag{2.5}$$

which obviously tends to zero as n tends to infinity.