Solutions to Sheet 2

Friedrich May, 355487; Markus Moll, 406263; Mariem Mounir, 415862

May 13, 2020

Exercise 1

a)

Let

$$X_2 = Y_1 Y_2$$

with Y_1, Y_2 representing one throw each. Because $\mathrm{E}(XY) = \mathrm{E}(X)\mathrm{E}(Y)$ for independant X, Y

$$\mathrm{E}(X_2) = \mathrm{E}(Y_1 Y_2) = \mathrm{E}(Y_1)\mathrm{E}(Y_2)$$

holds. With $\mathrm{E}(Y_1) = \mathrm{E}(Y_2) = 3.5$ this lead to

$$\mathrm{E}(X_2) = 3.5^2 = 12.25$$

. We further know that $\mathrm{Var}(X_2) = \mathrm{E}(X_2^2) - \mathrm{E}(X_2)^2$. Therefore

$$\mathrm{Var}(X_2) = \mathrm{E}(Y_1^2 Y_2^2) - 12.25^2 = \mathrm{E}(Y_1^2)\mathrm{E}(Y_2^2) - 12.25^2 = \frac{91^2}{36} - 12.25^2 \approx 79.97$$

, beacuse $\mathrm{E}(Y_1^2) = \mathrm{E}(Y_2^2) = \frac{91}{6}$.

b)

Let

$$X_n = Y_1 Y_2 \dots Y_n$$

with $Y_1, \dots, Y_n$ independent and describing one throw each. Then

$$\mathrm{E}(X_n) = \prod_{i=1}^{n} \mathrm{E}(Y_i)$$

. Using $\mathrm{E}(Y_i) = 3.5 \forall i$

$$\mathrm{E}(X_n) = 3.5^n$$

. As above

$$\mathrm{Var}(X_n) = \mathrm{E}(X_n^2) - \mathrm{E}(X_n)^2 = \prod_{i=1}^{n} \mathrm{E}(Y_i^2) - 3.5^{2n} = \frac{91^n - 73.5^n}{6^n}$$

.

c)

Let $X_i \in \{0;1\}$ with $X_i = \begin{cases} 1 & \text{throw results in one} \\ 0 & \text{otherwise} \end{cases}$ then

$$X = \sum_{i=1}^{200} X_i$$

is the number of ones in 200 throws. Observe that

$$\Pr(X_i = 1) = \frac{1}{6} = 1 - \Pr(X_i = 0) \forall i \in [n]$$

.

Then

$$\mathrm{E}(X) = \sum_{i=1}^{200} \mathrm{E}(X_i) = \sum_{i=1}^{200} \frac{1}{6} = \frac{200}{6}$$

$$\mathrm{E}(X^2) = \left(\sum_{i=1}^{200} \mathrm{E}(X_i) \right)^2 = \mathrm{E}(X)^2$$

. This leads to

$$\mathrm{Var}(X) = \mathrm{E}(X^2) - \mathrm{E}(X)^2 = 0$$

.

The Probaility to encounter at most 7 ones in 200 throws is

$$\Pr(X \leq 7) = \Pr(\text{"193 times not 1"}) = \prod_{i=1}^{193} \frac{5}{6} \leq 7 \times 10^{-16}$$

.

The inequalities yield the following borders:

Chebyshev

$$\Pr(|X - \mathrm{E}(X)| \geq \frac{158}{6}) \leq \frac{6^2 \mathrm{Var}(X)}{158} = 0$$

where $\frac{158}{6}$ is the minimum distance 7 ones are from the Expectation.

Chernoff

$$\Pr(X \leq \left(1 - \frac{158}{200}\right) \mathrm{E}(X)) \leq e^{-\frac{\frac{200}{6} \frac{158^2}{200^2}}{2}} = 0.0000303818$$

where $\left(1 - \frac{158}{200}\right) \mathrm{E}(X) = 7$.

Hoeffding

$$\Pr(X \leq \frac{200}{6} - \left(\frac{1}{6} - \frac{7}{200}\right) 200) \leq e^{-2\cdot 200\cdot\left(\frac{1}{6}-\frac{7}{200}\right)} = 0.009736638$$

In this case Chernoff's inequality gives the best bound. Chebyshev's inequality does not work here because $\mathrm{E}(X)^2 = \mathrm{E}(X^2) \Rightarrow \mathrm{Var}(X) = 0$.

Exercise 3

Using a description scheme where for each node is encoded using the binary representation of the index of the splitting feature using $\Sigma = \{0, 1\}$ and for each node $\Sigma^{\text{No. of features}}$ A 16 node decisiontree with 6 features uses $16 \cdot 2^3$ bits hence

$$n = 16 = 2^3 = 128.$$

We know from

$$m \geq \frac{1}{\epsilon}\left(n \ln|\Sigma| + 2\ln\left(\frac{2}{\delta}\right)\right)$$

that for $\epsilon = 0.05$ and $\delta = 0.2$ we need at least $m \geq 1821$ trianing examples to archive the described accuracy and propability for the described tree. Respecting the lower boud for m we can set $m \leq 1821$ as upper bound because 1821 training exaples surfice. Hence we don't need more than these samples.

Exercise 6

With $\alpha = \frac{1}{2}$, the event sequence 121234 and

$$L = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & \frac{1}{2} \end{pmatrix}$$

a)

Using the MWU algorithm, we find :

- $\omega^{(1)} = (1, 1, 1)$
- $\omega^{(2)} = (\frac{1}{2}, 1, \frac{1}{2})$
- $\omega^{(3)} = (\frac{1}{4}, \frac{1}{2}, \frac{1}{2})$
- $\omega^{(4)} = (\frac{1}{8}, \frac{1}{2}, \frac{1}{4})$
- $\omega^{(5)} = (\frac{1}{16}, \frac{1}{4}, \frac{1}{4})$
- $\omega^{(6)} = (\frac{1}{16}, \frac{1}{8}, \frac{1}{8})$
- And finally : $\omega^{(7)} = (\frac{1}{16}, \frac{1}{16}, 0.09)$

b)

Using the definition of the probability distribution, we find :

- The probability to follow the advice of expert 1 in round 6 is 0.2
- The probability to follow the advice of expert 2 in round 6 is 0.4
- The probability to follow the advice of expert 3 in round 6 is 0.4

So : $p^{(6)} = (0.2, 0.4, 0.4)$

c)

Even if we change the order the result of $\omega^{(7)}$ will stay the same, because it does not depend on the order :

$$\omega_i^{(t+1)} = (1-\alpha)^{\sum_{s=1}^{t} L_{ij(s)}} \omega_i^{(1)}$$