Statistical Machine Learning

Summer Term 2021, Homework 1

Prof. Stefan Roth, Dr. Simone Schaub-Meyer

TECHNISCHE UNIVERSITÄT DARMSTADT

Total points: ?? + ?? bonus
Due date: 25. May 2018

Stefanie Martin, Maximilian Nothnagel

Problem 1.1 Statistics Refreshner [0 Points]

a) Expectation & Variance [8 Points]

1. We can define the expectation by

$$E|f| = \sum_{w \in \Omega} P(w) f(w) \tag{18}$$

Which leads to the variance:

$$var[f] = E\left|f^2\right| - E[f]^2. \tag{19}$$

If we have 2 random variables X,Y and Z = X + Y. Then is the expectation a linear function, since for any 2 points

$$E[Z] = \sum_{w \in \Omega} Z(w) P(w) = \sum_{w \in \Omega} (X(w) + Y(w)) P(w) = E[X] + E[Y] \tag{20}$$

applies. Since the variance of the sum of 2 random variables is

$$var[Z] = E\left[Z^2\right] - E[Z]^2 = var|X| + var[Y] + 2E[XY] - 2E[X]E[Y] isE[XY] \neq E[X]E[Y] \tag{21}$$

$$Andthatisnotalinearoperator. \tag{22}$$

2. Unbiased estimator:

$$\overline{x} = \frac{1}{n} * \sum_{i=1}^{n} x_i \tag{23}$$

$$\overline{xA} = \frac{1}{6} * (1 + 5 + 6 + 3 + 2 + 1) = 3 \tag{24}$$

$$\overline{xB} = \frac{1}{6} * (6 + 1 + 1 + 4 + 1 + 5) = 3 \tag{25}$$

$$\overline{xC} = \frac{1}{6} * (3 + 2 + 3 + 3 + 4 + 3) = 3 \tag{26}$$

unbiased estimator for the variance

$$\overline{\sigma} = \frac{1}{n-1} * \sum_{i=1}^{n} x_i - \overline{x}^2 \tag{27}$$

$$\overline{\sigma A} = \frac{1}{5} * \left((1-3)^2 + (5-3)^2 + (6-3)^2 + (3-3)^2 + (2-3)^2 + (1-3)^2\right) = \frac{22}{5} = 4,4 \tag{28}$$

$$\overline{\sigma B} = \frac{1}{5} * \left((6-3)^2 + (1-3)^2 + (1-3)^2 + (4-3)^2 + (1-3)^2 + (5-3)^2\right) = \frac{26}{5} = 5,2 \tag{29}$$

$$\overline{\sigma C} = \frac{1}{5} * \left((3-3)^2 + (2-3)^2 + (3-3)^2 + (4-3)^2 + (3-3)^2 + (3-3)^2\right) = \frac{2}{5} = 0,4 \tag{30}$$

3.

$$KL : \sum_{x \in X} P(x) ln \frac{P(x)}{Q(x)} \tag{31}$$

$$KL(PA \parallel Q) = \frac{3}{6} * ln\left(\frac{\frac{3}{6}}{\frac{1}{6}}\right) + \frac{1}{6} * ln\left(\frac{\frac{1}{6}}{\frac{1}{6}}\right) + \frac{1}{6} * ln\left(\frac{\frac{1}{6}}{\frac{1}{6}}\right) + \frac{1}{6} * ln\left(\frac{\frac{1}{6}}{\frac{1}{6}}\right) + \frac{1}{6} * ln\left(\frac{\frac{1}{6}}{\frac{1}{6}}\right) = 3 \tag{32}$$

$$KL(PB \parallel Q) = \frac{1}{6} * ln\left(\frac{\frac{1}{6}}{\frac{1}{6}}\right) + \frac{2}{6} * ln\left(\frac{\frac{2}{6}}{\frac{1}{6}}\right) + \frac{1}{6} * ln\left(\frac{\frac{1}{6}}{\frac{1}{6}}\right) + \frac{1}{6} * ln\left(\frac{\frac{1}{6}}{\frac{1}{6}}\right) = 1,52 \tag{33}$$

$$KL(PC \parallel Q) = \frac{4}{6} * ln\left(\frac{\frac{4}{6}}{\frac{1}{6}}\right) + \frac{1}{6} * ln\left(\frac{\frac{1}{6}}{\frac{1}{6}}\right) + \frac{1}{6} * ln\left(\frac{\frac{1}{6}}{\frac{1}{6}}\right) = 2.38 \tag{34}$$

A has the biggest KL divergence, so it is the closest

b) It is a cold world [7 Points]

1.

$$a \in \{0,1\} : if apersonhasbackspin, with 1 = painand 0 = nopain \tag{43}$$

$$b \in \{0,1\} : if apersonhasacold, with 1 = painand 0 = nocold \tag{44}$$

2.

$$P(a=1 \mid b=1) = 0,25 \tag{45}$$

$$P(b=1) = 0,04 \tag{46}$$

$$P(a=1 \mid b=0) = 0,1 \tag{47}$$

3. Rule of Bayes

$$P(b=1 \mid a=1) = \frac{P(a=1 \mid b=1) P(b=1)}{P(b=1)} \tag{48}$$

$$\frac{P(a=1 \mid b=1) P(b=1)}{P(a=1 \mid b=1) P(b=1) + P(a=1 \mid b=0) P(b=0)} \tag{49}$$

$$Werteeinsetzen : \frac{0,25 * 0,04}{0,25 * 0,04 + 0,10 * (1-0,04)} = \frac{5}{53} \approx 0,094 \tag{50}$$

c) Cure the virus [14 Points]

3. Markov Chain

$$S_0 = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \tag{56}$$

$$S_1 = \begin{pmatrix} 0,42 \\ 0,58 \end{pmatrix} \tag{57}$$

(58)

(59)

(60)

2.
```
import numpy as np
import matplotlib.pyplot as plt

def markovChain(s0, p, g):
    s = s0.copy()
    result = np.zeros(g+1)
    result[0]=s[0]
    for i in range(1, g+1):
        s = s.dot(p)
        result[i]=s[0]
    return result

s0 = np.array([1, 0])
s1 = np.array([0.026, 0.974])
p = np.array([[0.42, 0.58],[0.026, 0.974]])
n = np.arange(0, 19, 1)

gen18 = markovChain(s0, p, 18)
gen18prog = markovChain(s1, p, 18)

plt.plot(n, gen18, 'b')
plt.plot(n, gen18prog, 'r')
plt.xlabel('Generations')
plt.ylabel('Probability')
plt.title('Virus through the generations Task 3')
plt.grid(True)
plt.show()
```

Problem 1.2 Information Theory [0 Points]

a) Entropy [5 Points]

1.

$$-0,04 * \log_2 0,04 - 0,22 * \log_2 0,22 - 0,67 * \log_2 0,67 - 0,07 * \log_2 0,07 = 1,3219 \tag{62}$$

2.