Name:
NetID:

COS 302 – Final Practice Questions

Problem 1

Two counterfeit coins of equal weight are mixed with 8 identical genuine coins. The weight of each of the counterfeit coins is different from the weight of each of the genuine coins. A pair of coins is selected at random without replacement from the 10 coins. A second pair is selected at random without replacement from the remaining 8 coins. The combined weight of the first pair is equal to the combined weight of the second pair. What is the probability that all 4 selected coins are genuine?

(A) $\frac{7}{11}$

(B) $\frac{9}{13}$

(C) $\frac{11}{15}$

(D) $\frac{15}{19}$

Note that we are trying to find the conditional probability $P(A|B) = \frac{P(A\cap B)}{P(B)}$ where A is the 4 coins being genuine and B is the sum of the weight of the coins being equal. The only possibilities for B are (g is abbreviated for genuine and c is abbreviated for counterfeit, and the pairs are the first and last two in the quadruple) $(g, g, g, g), (g, c, g, c), (g, c, c, g), (c, g, g, c), (c, g, c, g)$. We see that $A \cap B$ happens with probability $\frac{8}{10} \times \frac{7}{9} \times \frac{6}{8} \times \frac{5}{7} = \frac{1}{3}$, and B happens with probability $\frac{1}{3} + 4 \times \left(\frac{8}{10} \times \frac{2}{9} \times \frac{7}{8} \times \frac{1}{7}\right) = \frac{19}{45}$, hence $P(A|B) = \frac{P(A\cap B)}{P(B)} = \frac{\frac{1}{3}}{\frac{19}{45}} = \boxed{\frac{15}{19} \text{ } (\mathbf{D})}$.

Problem 2

You flip a biased coin to decide which of two books to study. With probability $1/3$ you study a chapter of linear algebra, and with probability $2/3$ you study a chapter of a book on machine learning. The number of typos in a chapter of linear algebra textbook follows a Poisson distribution with $\lambda = 1$. The number of typos in a chapter of machine learning textbook follows a Poisson distribution with $\lambda = 2$. What is the expected number of typos you will encounter?

(Poisson distribution with parameter λ has its pmf given by $p(x = k) = \frac{\lambda^k e^{-\lambda}}{k!}$ and its mean equal to λ.)

(A) $5/3$

(B) $2/3$

(C) 1

(D) $1/3$

Answer: $\boxed{\textbf{(A)}}$

As given above, λ is the expectation (mean) of the Poisson distribution. The expected number of typos in a chapter of linear algebra textbook is 1 and the expected number of typos in a chapter of machine learning textbook is 2. The expected number of typos is $1/3 * 1 + 2/3 * 2 = 5/3$.

Problem 3

Suppose W_1, W_2, W_3 is a collection of random variables. Suppose also that W_2 and W_3 are independent, but that W_1 might depend on one of the others. Assuming all expectations exist, which of the following statements are always true?

(A) $\mathbb{E}[W_1 + W_2] = \mathbb{E}[W_1] + \mathbb{E}[W_2]$

(B) $\mathrm{Var}[W_1 + W_2] = \mathrm{Var}[W_1] + \mathrm{Var}[W_2]$

(C) $\mathrm{Var}[W_2 W_3] = \mathrm{Var}[W_2]\,\mathrm{Var}[W_3]$

(D) $\mathbb{E}[W_2 W_3] = \mathbb{E}[W_2]\mathbb{E}[W_3]$

(E) $\mathbb{E}[W_1 W_2 W_3] = \mathbb{E}[W_1]\mathbb{E}[W_2]\mathbb{E}[W_3]$

(F) For any $a, b, c \in \mathbb{R}$, $\mathrm{Var}[a^3 b W_2 + b^4 a W_3 + c] = a^6 b^2\,\mathrm{Var}[W_2] + b^8 a^2\,\mathrm{Var}[W_3]$

Answer:

(A) True

(B) False: only true when they are independent

(C) False

(D) True

(E) False

(F) True

Problem 4

Let X and Y be two (not necessarily independent) random variables both with mean 0 and variance 1. Let $Z = X + Y$. Select all statements that are not possible.

(A) $\mathbb{E}[Z] = 0$

(B) $\mathrm{Var}[Z] = 0$

(C) $\mathbb{E}[Z] = 1$

(D) $\mathrm{Var}[Z] = 2$

Answer: $\boxed{\textbf{only (C)}}$. By linearity of expectation, $\mathbb{E}[Z] = \mathbb{E}[X] + \mathbb{E}[Y] = 0$. Also note that $\mathrm{Var}[Z] = 0$ when $Y = -X$, and $\mathrm{Var}[Z] = 2$ when X and Y are independent.

Problem 5

Which of the following statements is **not equivalent** to the statement "function $f(\mathbf{x})$ is convex"?

(A) For any two points $\mathbf{x}, \mathbf{y}$ it holds that $f(\mathbf{y}) \geq f(\mathbf{x}) + \nabla_x f(\mathbf{x})(\mathbf{y} - \mathbf{x})$.

(B) $\nabla_x^2 f(\mathbf{x})$ is positive semi-definite (suppose $f(\mathbf{x})$ is twice differentiable).

(C) For any two points $\mathbf{x}, \mathbf{y}$ and $\theta \in [0, 1]$ it holds that $f(\theta\mathbf{x} + (1 - \theta)\mathbf{y}) \leq \theta f(\mathbf{x}) + (1 - \theta) f(\mathbf{y})$.

(D) A global minimum exists for function $f(\mathbf{x})$.

Answer: D

Convex functions might be monotonic, and not have any minima. For example, consider $f(x) = e^x$.

Problem 6

Let X_i for $i = 1, 2, \ldots, 1{,}000{,}000$ be i.i.d. variables with mean and variance equal to 1 (for example, they might be Poisson-distributed variables with $\lambda = 1$). Define $S_n = \sum_{i=1}^{n} X_i$. Recall that the Central Limit Theorem states that for i.i.d. random variables X_i with mean μ and variance σ^2, the sequence $\sqrt{n}\,(\overline{X}_n - \mu)$, where $\overline{X}_n = S_n/n$ converges in distribution to a normal distribution with mean 0 and variance σ^2. Using the CLT, estimate, to the nearest whole-number percentage,

$$P(998000 \leq S_{1000000} \leq 1002000).$$

The idea is to use the CLT; the variance of each X_i is 1, and since they are independent, $\mathrm{Var}(S_{1000000}) = 1000000$. This implies that the standard deviation of $S_{1000000}$ is 1000. And since the expectation of X_i is 1, we are estimating the desired quantity by the probability that a Gaussian lies within two standard deviations of the mean, which is about 95%.

Problem 7

We're going to implement gradient descent by hand. Recall the gradient descent update rule:

$$x_{t+1} = x_t - \lambda(\nabla_x f(x_t))^T,$$

where λ is called the learning rate.

Consider the function $f(x, y) = 4xy$. Suppose we initialize our starting point at $(5, 5)$. Now, using a learning rate of 0.1, what will be the coordinates (x, y) of the point we will end up at after 2 steps?

(A) $[0, 0]$

(B) $[9.8, 9.8]$

(C) $[1.8, 1.8]$

(D) $[-1.8, 1.8]$

Answer: C

The gradient of the function is $(4y, 4x)$. Then, we can write the update rules

$$\begin{bmatrix} 5 \\ 5 \end{bmatrix} - 0.1 \begin{bmatrix} 20 \\ 20 \end{bmatrix} = \begin{bmatrix} 3 \\ 3 \end{bmatrix}$$

$$\begin{bmatrix} 3 \\ 3 \end{bmatrix} - 0.1 \begin{bmatrix} 12 \\ 12 \end{bmatrix} = \begin{bmatrix} 1.8 \\ 1.8 \end{bmatrix}$$

Problem 8

Fill in the blank, using “sometimes,” “always,” or “never”. Suppose we are running gradient descent. Then regardless of the point at which we initialize our descent algorithm, if we choose a small enough learning rate, the algorithm will __________________ converge to the global minimum.

Sometimes. This is because it’s possible for the gradient descent to get "stuck" at local minima (unless we have a convex function), depending on where we initialize the descent.

Problem 9

Suppose we have four lilypads in a row (labeled 1, 2, 3, 4, 5, in that order), and Kenny the frog is sitting on lilypad number 2. At each step, Kenny will jump left or right 1 lilypad with probability 1/2, independently of other jumps. Unfortunately, there is an alligator sitting on the 5th lilypad, and if Kenny jumps to lilypad 5, he will be eaten. If Kenny makes it to lilypad 1, where his home is located, then he will stay there. What is the probability that Kenny will make it home without being eaten?

(A) $\frac{3}{4}$

(B) $\frac{1}{2}$

(C) $\frac{4}{5}$

(D) $\frac{1}{4}$

We can define P_i to be the probability that Kenny makes it home, given that it is on lilypad i. We have that $P_1 = 1, P_5 = 0$, since if Kenny is on lilypad 1, then he makes it home with probability 1, and if he lands on lilypad 5, then he will not make it home. Then we have the equations $P_2 = \frac{1}{2} \cdot 1 + \frac{1}{2}P_3$, $P_3 = \frac{1}{2} \cdot P_2 + \frac{1}{2} \cdot P_4$, $P_4 = \frac{1}{2}P_3$. Solving yields $P_2 = \frac{3}{4}$.

Problem 10

X and Y are uncorrelated random variables (i.e., their correlation is zero). Which of the following must be true:

(A) The covariance between X and Y is zero.

(B) X and Y are independent.

(C) Both of the above.

(D) Neither of the above.

Solution: (A)

Problem 11

Consider univariate functions of the form:

$$f(x) = ax^3 + bx^2 + cx + d,$$

with coefficients $a, b, c, d \neq 0$ and input $x \in \mathbb{R}$. The function `get_stationary` below is supposed to return a list of the unique x coordinates of all real stationary points of the function $f(x)$ defined by the provided coefficients. Fill in the blanks in the following code snippet to complete the implementation.

```
import numpy as np

def get_stationary(a, b, c, d):
    if 0 in [a, b, c, d]:
        raise ValueError("Only non-zero coefficients are accepted.")
    a_new = 3*a
    b_new = BLANK_1
    c_new = BLANK_2
    coords = []
    sqrt_term = BLANK_3 - (4*a_new*c_new)
    if sqrt_term < 0:
        return coords
    sqrts = [np.sqrt(sqrt_term)]
    if sqrt_term != 0:
        sqrts.append(-1*sqrts[0])
    for s in sqrts:
        coords.append((-b_new + s) / (BLANK_4))
    return coords
```

Answers:
BLANK_1: `2*b`
BLANK_2: `c`
BLANK_3: `b_new**2`
BLANK_4: `2*a_new`

Problem 12
Which of these is the Jacobian of $\boldsymbol{f}(\boldsymbol{x}) = \operatorname{tr}(\boldsymbol{x}\boldsymbol{x}^\top \boldsymbol{A})$ where $\boldsymbol{x} \in \mathbb{R}^d$ and $\boldsymbol{A} \in \mathbb{R}^{d\times d}$?

(A) $\boldsymbol{x}^\top(\boldsymbol{A} + \boldsymbol{A}^\top)$

(B) $\boldsymbol{A}\boldsymbol{x}$

(C) $\operatorname{tr}(\boldsymbol{x}^\top \boldsymbol{A})$

(D) $\det(\boldsymbol{x}\boldsymbol{x}^\top \boldsymbol{A})$

(E) None of the above.

$$\begin{aligned}
\boldsymbol{f}(\boldsymbol{x}) &= \operatorname{tr}(\boldsymbol{x}\boldsymbol{x}^\top \boldsymbol{A}) \\
&= \operatorname{tr}(\boldsymbol{A}\boldsymbol{x}\boldsymbol{x}^\top) \\
&= \operatorname{tr}(\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x}) \\
&= \boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x}
\end{aligned}$$

(A) Using standard identity for quadratics, the Jacobian is $\boldsymbol{x}^\top(\boldsymbol{A} + \boldsymbol{A}^\top)$.

Problem 13

Consider a unit square with vertices A, B, C, and D. Choose a point P uniformly at random inside the square. Let $d^2(X, Y)$ denote the squared distance between the points X and Y. Find the expected value of $d^2(A, P) + d^2(B, P) + d^2(C, P) + d^2(D, P)$. (Hint: let $A = (0, 0)$, $B = (1, 0)$, $C = (1, 1)$, and $D = (0, 1)$.)

(A) 2

(B) 8/3

(C) 3/2

(D) 16/9

Answer: $\boxed{\textbf{(B)}}$. First notice that by linearity of expectation and by symmetry,

$$\begin{aligned}\mathbb{E}[d^2(A, P) + d^2(B, P) + d^2(C, P) + d^2(D, P)] &= \mathbb{E}[d^2(A, P)] + \mathbb{E}[d^2(B, P)] + \mathbb{E}[d^2(C, P)] + \mathbb{E}[d^2(D, P)] \\ &= 4\mathbb{E}[d^2(A, P)].\end{aligned}$$

We can let $A = (0, 0)$, $B = (1, 0)$, $C = (1, 1)$, and $D = (0, 1)$. If $P = (x, y)$, then $d^2(A, P) = x^2 + y^2$. Since P is uniformly distributed inside the square, x and y are each uniformly distributed in the interval $(0, 1)$. So

$$\mathbb{E}[d^2(A, P)] = \int_0^1 \int_0^1 (x^2 + y^2)\, dx\, dy = 2/3.$$

Multiplying by 4, we get 8/3.