18-661: Introduction to ML for Engineers

Math Quiz Review

Spring 2025

ECE – Carnegie Mellon University

Topics

Not an exhaustive list . . .

- Basic probability; conditional probability; Bayes' theorem
- Correlation, independence
- Continuous & discrete random variables; PDFs & PMFs; conditional PDFs and PMFs
- Expectation, variance
- Common distributions (e.g. Bernoulli, Binomial/Poisson, Geometric/Exponential, . . .), Normal/Multivariate Normal
- MLE/MAP

But if you aren't familiar with most of these things, you should review on your own ASAP

Quiz

True/false questions:

- If your answer is sometimes, not always, etc, be prepared to come up with a counter-example
- If your answer is always, never, etc, be prepared to state your reasoning

Not actually a quiz

. . . but please do try, since you'll be in for a rough time if your answers aren't *confident and correct* most of the time and don't realize until later in the course.

Q1: Basic Probability

Which are true (and how can you fix the ones that are false)?

(a) $P(A \cup B) = P(A) + P(B)$

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

(c) $P(A) = 1 - P(A^c)$ or $P(A) = 1 - P(\bar{A})$

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

Q2: Probability of CMU Graduate Student

Problem: Find the probability that a graduate student is studying at CMU, given that the student is sleep-deprived.

Given:

- $a\%$ of all graduate students attend CMU.
- $b\%$ of CMU students are sleep-deprived.
- $\frac{b}{200}\%$ of students at other colleges are sleep-deprived.

Options:

$$\text{(A)} \quad \frac{a}{a + \frac{100-a}{200}}$$

$$\text{(B)} \quad \frac{b}{a + \frac{100-b}{200}}$$

$$\text{(C)} \quad \frac{a}{b + \frac{100-a}{200}}$$

$$\text{(D)} \quad \frac{b}{b + \frac{100-b}{200}}$$

Q2: Probability of CMU Graduate Student

Key Concept: Bayes' Theorem

$$P(\text{CMU} \mid \text{Sleep-Deprived}) = \frac{P(\text{Sleep-Deprived} \mid \text{CMU}) \cdot P(\text{CMU})}{P(\text{Sleep-Deprived})}$$

Step-by-Step Guide:

- Calculate $P(\text{Sleep-Deprived} \mid \text{CMU}) = \frac{b}{100}$.
- Calculate $P(\text{Sleep-Deprived} \mid \text{Other Colleges}) = \frac{b}{20000}$.
- Use the total probability formula to compute $P(\text{Sleep-Deprived})$:

$$P(\text{Sleep-Deprived}) =$$

$$P(\text{Sleep-Deprived} \mid \text{CMU})P(\text{CMU})$$

$$+P(\text{Sleep-Deprived} \mid \text{Other Colleges})P(\text{Not CMU}).$$

- Plug values into Bayes' Theorem.

Q3: Expectation and Variance

Problem: Consider a continuous random variable x that is uniformly distributed between 2 and 4. What are the mean and variance of x?

Options:

1. $3, \frac{\sqrt{3}}{3}$
2. $0, \frac{\sqrt{3}}{3}$
3. $3, \frac{1}{3}$
4. $3, 1$

Q3: Expectation and Variance

Key Concept: Uniform Distribution A continuous uniform random variable $x \sim U(a, b)$ has:

- **Mean:** $\mu = \frac{a+b}{2}$
- **Variance:** $\sigma^2 = \frac{(b-a)^2}{12}$

Solution Steps:

- Here, $a = 2$ and $b = 4$.
- Calculate the mean:
$$\mu = \frac{2+4}{2} = 3$$
- Calculate the variance:
$$\sigma^2 = \frac{(4-2)^2}{12} = \frac{4}{12} = \frac{1}{3}$$

Q3: Expectation and Variance

Is it possible for a random variable to have . . .

(a) Infinite mean and infinite variance?

(b) Finite mean and infinite variance?

(c) Infinite mean and finite variance?

Side-note: what would it look like if you were sampling from a random variable with an infinite mean?

Side-side-note: do random variables with infinite mean exist in the real world?

Q4: Matrix Algebra

Problem: Which one of the following identities is **incorrect** for two arbitrary real-valued matrices A and B? Assume that the inverses exist and the multiplications are valid.

Options:

1. $(AB)^\top = B^\top A^\top$
2. $(A+B)^{-1} = A^{-1} + B^{-1}$
3. $(A^\top)^{-1} = (A^{-1})^\top$
4. $\det(A^{-1}) = \det(A)^{-1}$

Q4: Matrix Algebra

Key Concepts:

- **Matrix Transpose Rules:** $(AB)^\top = B^\top A^\top$ is valid.
- **Matrix Inversion:** $(A+B)^{-1} \neq A^{-1} + B^{-1}$. The inverse of a sum cannot be expressed as the sum of inverses.
- **Inverse of a Transpose:** $(A^\top)^{-1} = (A^{-1})^\top$ is valid.
- **Determinant of an Inverse:** $\det(A^{-1}) = \det(A)^{-1}$ is valid.

Q5: Convexity

Definition: A function $f(x)$ is convex if, for all $x, y \in \text{domain}(f)$ and $t \in [0, 1]$:

$$f(tx + (1-t)y) \leq tf(x) + (1-t)f(y).$$

Key Properties of Convex Functions:

- The second derivative (Hessian) is positive semi-definite ($H \succeq 0$).
- Linear and exponential functions are convex.
- The sum of convex functions is also convex.

Non-Convexity: A function is non-convex if it violates the above inequality or its Hessian is not positive semi-definite.

Q5: Convexity

Step-by-Step Method:

1. Compute the first and second derivatives (if the function is differentiable).
2. Check the Hessian matrix ($\nabla^2 f(x)$):
   - If $H \succeq 0$, the function is convex.
   - If $H \not\succeq 0$, the function is not convex.
3. For non-differentiable functions, use the definition directly:

$$f(tx + (1-t)y) \leq tf(x) + (1-t)f(y).$$

Examples:

- $\sum_{i=1}^{n} a_i x_i$ is convex (linear function).
- $\sum_{i=1}^{n} \exp(x_i + a_i)$ is convex (exponential function).

Q5: Convexity

Problem: Determine which function is NOT convex:

$$f(x) = \sum_{i=1}^{n} a_i x_i$$

$$f(x) = \sum_{i=1}^{n} a_i |x_i|$$

$$f(x) = \sum_{i=1}^{n} \exp(x_i + a_i)$$

$$f(x) = \sum_{i=1}^{n} (a_i - \log(x_i))$$

Q5: Convexity

Problem: Determine which function is NOT convex:

$$f(x) = \sum_{i=1}^{n} a_i x_i \quad \text{(Convex, linear function)}$$

$$f(x) = \sum_{i=1}^{n} a_i |x_i| \quad (\textbf{Not Convex}, \text{ as } |x| \text{ is not linear})$$

$$f(x) = \sum_{i=1}^{n} \exp(x_i + a_i) \quad \text{(Convex, exponential)}$$

$$f(x) = \sum_{i=1}^{n} (a_i - \log(x_i)) \quad \text{(Convex, logarithmic term is concave in domain.)}$$

Answer: The function $f(x) = \sum_{i=1}^{n} a_i |x_i|$ is NOT convex because the absolute value function ($|x_i|$) is piecewise linear and not smooth.

Q6 Jensen's Inequality

Jensen's Inequality: For a convex function g, the inequality states:

$$\mathbb{E}[g(X)] \geq g(\mathbb{E}[X]),$$

where X is a random variable and $\mathbb{E}[X]$ is its expected value.

Key Notes:

- If g is concave, the inequality reverses:

$$\mathbb{E}[g(X)] \leq g(\mathbb{E}[X]).$$

- Convex functions "curve upwards," while concave functions "curve downwards."
- Jensen's inequality is foundational in probability and optimization.

Examples:

- Convex: $g(x) = x^2$, $g(x) = \exp(x)$.
- Concave: $g(x) = \log(x)$, $g(x) = \sqrt{x}$.

Q6 Jensen's Inequality

Problem: Which statements are true based on Jensen's inequality?

Analysis:

- $\mathbb{E}[g(X)] \geq g(\mathbb{E}[X])$ for a convex function g
- $\mathbb{E}[g(X)] \geq g(\mathbb{E}[X])$ for a concave function g
- $\log\left(\frac{x_1+x_2}{2}\right) \geq \frac{1}{2}(\log x_1 + \log x_2)$
- $\exp\left(\sum_{i=1}^{n} \alpha_i x_i\right) \geq \sum_{i=1}^{n} \alpha_i \exp(x_i)$, where $\alpha_i \geq 0$ and $\sum_{i=1}^{n} \alpha_i = 1$

Q6 Jensen's Inequality

Problem: Which statements are true based on Jensen's inequality?

Analysis:

- $\mathbb{E}[g(X)] \geq g(\mathbb{E}[X])$ for a convex function g:
  True, as this is the direct statement of Jensen's inequality.
- $\mathbb{E}[g(X)] \geq g(\mathbb{E}[X])$ for a concave function g:
  False, as the inequality reverses for concave functions.
- $\log\left(\frac{x_1+x_2}{2}\right) \geq \frac{1}{2}(\log x_1 + \log x_2)$:
  True, since $\log(x)$ is concave, and the inequality follows Jensen's reversed form.
- $\exp\left(\sum_{i=1}^{n} \alpha_i x_i\right) \geq \sum_{i=1}^{n} \alpha_i \exp(x_i)$, where $\alpha_i \geq 0$ and $\sum_{i=1}^{n} \alpha_i = 1$:
  False, as $\exp(x)$ is convex, and the correct inequality is:

$$\mathbb{E}[\exp(X)] \geq \exp(\mathbb{E}[X]).$$

Q7: Calculus and Q8: Matrix Calculus

Problem 7: Derivative of $f(w) = l(w) + \lambda w^2$

Given: $l(w)$ is a differentiable function and λ is a constant.

Solution: Using basic derivative rules:

$$\frac{\partial f(w)}{\partial w} = \frac{\partial l(w)}{\partial w} + \frac{\partial}{\partial w}(\lambda w^2).$$

$$\frac{\partial f(w)}{\partial w} = l'(w) + 2\lambda w.$$

Problem 8: Derivative of $\alpha = \mathbf{x}^\top A\mathbf{x}$

Given: A is a symmetric matrix, and $\mathbf{x} \in \mathbb{R}^n$.

Solution: Using matrix calculus:

$$\frac{\partial}{\partial \mathbf{x}}\left(\mathbf{x}^\top A\mathbf{x}\right) = 2A\mathbf{x} \text{or} 2A^\top \mathbf{x}.$$

Correct Answer: $2A\mathbf{x}$ or $2A^\top \mathbf{x}$.

Q9 Eigenvalues

Problem: Suppose matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ is such that $\text{rank}(\mathbf{A}) = n$. Then all the eigenvalues of $\mathbf{A}$ are:

- Positive
- Negative
- Non-zero
- Zero

Q9 Eigenvalues

Problem: Suppose matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ is such that $\text{rank}(\mathbf{A}) = n$. Then all the eigenvalues of $\mathbf{A}$ are:

- Positive
- Negative
- **Non-zero (Correct Answer)**
- Zero

Solution:
1. $\text{rank}(\mathbf{A}) = n$ implies $\mathbf{A}$ is **full rank**, meaning all rows (or columns) are linearly independent.
2. For a square $n \times n$ matrix:
- The rank equals the number of non-zero eigenvalues.
- Full rank ($\text{rank} = n$) implies all n eigenvalues are **non-zero**.
3. If any eigenvalue were zero, $\mathbf{A}$ would lose full rank.

Conclusion: Since $\text{rank}(\mathbf{A}) = n$, all eigenvalues of $\mathbf{A}$ are **non-zero**.

Q10 Eigenvalues

Problem: Suppose $\mathbf{A} \in \mathbb{R}^{n \times n}$ is a positive semi-definite matrix. Which of the following statements are true?

1. $\mathbf{A}$ is full rank.
2. All the eigenvalues of $\mathbf{A}$ are non-negative.
3. All the eigenvalues of $\mathbf{A}$ are positive.
4. $\mathbf{x}^\top \mathbf{A} \mathbf{x} > 0$ for all $\mathbf{x} \in \mathbb{R}^n$.
5. $\mathbf{x}^\top \mathbf{A} \mathbf{x} \geq 0$ for all $\mathbf{x} \in \mathbb{R}^n$.

Q10 Eigenvalues

Problem: Suppose $\mathbf{A} \in \mathbb{R}^{n\times n}$ is a positive semi-definite matrix. Which of the following statements are true?

1. $\mathbf{A}$ is full rank. Not necessarily true. A positive semi-definite matrix can have eigenvalues equal to zero, meaning it may not be full rank.
2. All the eigenvalues of $\mathbf{A}$ are non-negative. **True**. By definition, a positive semi-definite matrix has eigenvalues ≥ 0.
3. All the eigenvalues of $\mathbf{A}$ are positive. False. This would be true for positive definite matrices, not for semi-definite ones.
4. $\mathbf{x}^\top \mathbf{A}\mathbf{x} > 0$ for all $\mathbf{x} \in \mathbb{R}^n$. False. This holds for positive definite matrices, not semi-definite ones (equality may occur for some $\mathbf{x}$).
5. $\mathbf{x}^\top \mathbf{A}\mathbf{x} \geq 0$ for all $\mathbf{x} \in \mathbb{R}^n$. **True**. By definition, positive semi-definite matrices satisfy this inequality.

Q11 Geometry

Problem 1: Geometric Interpretation of $\mathbf{w}^\top \mathbf{x} + b = 0$

- $\mathbf{w} \in \mathbb{R}^3$ and $b \in \mathbb{R}$ are constants.
- The equation $\mathbf{w}^\top \mathbf{x} + b = 0$ represents a **plane** in $\mathbb{R}^3$.
- Reason: A plane is defined as the set of points $\mathbf{x}$ that satisfy a linear equation with constant coefficients.

Problem 2: Perpendicular Vector to $\mathbf{w}$

- Given two vectors $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^3$ satisfying $\mathbf{w}^\top \mathbf{x} + b = 0$, the vector:

$$\mathbf{x}_1 - \mathbf{x}_2 \quad \text{is perpendicular to } \mathbf{w}.$$

- Reason: Subtracting two points on the plane gives a direction vector on the plane. The normal vector $\mathbf{w}$ is orthogonal to the plane.

Problem 3: Shortest Distance Between Two Parallel Planes

- Consider the planes $\mathbf{w}^\top \mathbf{x} + b_1 = 0$ and $\mathbf{w}^\top \mathbf{x} + b_2 = 0$.
- The shortest distance between the planes is given by:

$$\frac{|b_1 - b_2|}{\sqrt{\mathbf{w}^\top \mathbf{w}}}.$$

- Reason: The distance between two parallel planes is the perpendicular distance, which depends on the offset $|b_1 - b_2|$ and the magnitude of $\mathbf{w}$.

Q12 Vector-Vector Multiply

Problem: Given the two mathematical expressions:

$$\text{(i)} \quad \begin{bmatrix} 5 \\ -2 \end{bmatrix}^\top \begin{bmatrix} 5 \\ -2 \end{bmatrix}, \quad \text{(ii)} \quad \begin{bmatrix} 5 \\ -2 \end{bmatrix} \begin{bmatrix} 5 \\ -2 \end{bmatrix}^\top,$$

determine which of the statements are true.

Solution:

- **Expression (i): Inner Product**
  - The operation $\mathbf{v}^\top\mathbf{v}$ results in a scalar:

$$\begin{bmatrix} 5 & -2 \end{bmatrix} \begin{bmatrix} 5 \\ -2 \end{bmatrix} = 25 + 4 = 29.$$

  - This is an **inner product**, producing a scalar.
- **Expression (ii): Outer Product**
  - The operation $\mathbf{v}\mathbf{v}^\top$ results in a 2×2 matrix:

$$\begin{bmatrix} 5 \\ -2 \end{bmatrix} \begin{bmatrix} 5 & -2 \end{bmatrix} = \begin{bmatrix} 25 & -10 \\ -10 & 4 \end{bmatrix}.$$

  - This is an **outer product**, producing a matrix.

Q13 Singular Value Decomposition

Problem: Given the SVD of a matrix $A \in \mathbb{R}^{m \times n}$:

$$A = U\Sigma V^\top,$$

which of the following is equivalent to $A^\top A$?

Solution: 1. Start with the definition of $A^\top A$:

$$A^\top A = (U\Sigma V^\top)^\top (U\Sigma V^\top).$$

2. Use the transpose property $(AB)^\top = B^\top A^\top$:

$$A^\top A = (V\Sigma^\top U^\top)(U\Sigma V^\top).$$

3. Simplify using the orthogonality of U ($U^\top U = I$):

$$A^\top A = V\Sigma^\top \Sigma V^\top.$$

4. Recognize that $\Sigma^\top \Sigma$ is a diagonal matrix with squared singular values (Σ^2):

$$A^\top A = V\Sigma^2 V^\top.$$

Q14 Probability Distributions

Problem: Match each distribution to its corresponding probability density function (PDF):

Distribution	Probability Density Function (PDF)
(1) Normal	(a) $\binom{n}{k} p^k (1-p)^{n-k}$
(2) Binomial	(b) $\frac{1}{b-a}$ when $a \leq x \leq b$; 0 otherwise
(3) Uniform	(c) $\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)$
(4) Bernoulli	(d) $p^x (1-p)^{1-x}$

Solution:

- (1) Normal $\rightarrow$ (c): $\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)$ **Reason:** This is the PDF of a standard normal distribution.
- (2) Binomial $\rightarrow$ (a): $\binom{n}{k} p^k (1-p)^{n-k}$ **Reason:** The binomial distribution gives the probability of k successes in n trials.
- (3) Uniform $\rightarrow$ (b): $\frac{1}{b-a}$ when $a \leq x \leq b$; 0 otherwise **Reason:** The uniform distribution is constant between a and b.
- (4) Bernoulli $\rightarrow$ (d): $p^x (1-p)^{1-x}$ **Reason:** The Bernoulli distribution models a single trial with success probability p.

Q14 Probability Distributions

Would it be correct to do the following:

(a) Model the average of a very large number of Exponential random variables using a Gaussian random variable

$$\frac{1}{\sqrt{2\pi\sigma^2}} \exp \frac{(x-\mu)^2}{-2\sigma^2}$$

(b) Model the time (in minutes) between visitors to a Coffee shop using a Geometric random variable

$$P(X = k) = (1-p)^{k-1} p$$

how about

$$\lambda \exp^{-\lambda x} \quad (x \geq 0)$$

?

(c) Model the number of visitors to a Coffee shop during a period of time using a Poisson random variable

$$\frac{\lambda^k e^{-\lambda}}{k!}$$

Q15 Python

Problem: Analyze the Python class and determine outputs for the given instructions. Code Summary: Class ‘stackofnums’ implements a stack with the following methods:

- ‘push(x)’: Adds an element x to the stack.
- ‘pop()’: Removes the top element of the stack if not empty.
- ‘top()’: Returns the top element of the stack if not empty.
- ‘average()‘: Computes the average of all elements in the stack.

Problem 1:

$$s = \text{stackofnums}()$$
$$s.\text{push}(2)$$
$$s.\text{push}(7)$$
$$s.\text{pop}() \quad \text{(removes 7)}$$
$$s.\text{push}(8)$$
$$s.\text{top}() \quad \text{(returns 8)}.$$

Output: 8.

Q15 Python

Problem: Analyze the Python class and determine outputs for the given instructions. Code Summary: Class ‘stackofnums’ implements a stack with the following methods:

- ‘push(x)’: Adds an element x to the stack.
- ‘pop()’: Removes the top element of the stack if not empty.
- ‘top()’: Returns the top element of the stack if not empty.
- ‘average()‘: Computes the average of all elements in the stack.

Problem 2:

$$s.\text{pop}() \quad \text{(removes 8, stack: [2])}$$
$$s.\text{average}() \quad \text{(average of [2]).}$$

Output: 2.