CS 181 and CSCI E-181 Calibration Questions

This is a short ungraded quiz, which will help you calibrate your background for the course. These questions are not in any sense a comprehensive list of topics that you need to know for the course. Rather, you should think of these as being "representative" of topics with which you should be comfortable. If you find these questions impossibly difficult or using language that you find unfamiliar, then it is strongly suggested that you wait to take CS181 until you have covered more of the prerequisites. When you fill out the course survey, please include your score. Answers are provided.

1. **The Gaussian/Normal Distribution**

   Which of the following is the PDF of the Gaussian (or normal) distribution?

   (A) $\mathcal{N}(x \mid \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$

   (B) $\mathcal{N}(x \mid \mu, \sigma^2) = \frac{\mu^\sigma}{\Gamma(\sigma)} x^{\sigma-1} \exp\{-x\mu\}$

   (C) $\mathcal{N}(x \mid \mu, \sigma^2) = \frac{x}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(\ln x-\mu)^2}{2\sigma^2}\right\}$

   (D) $\mathcal{N}(x \mid \mu, \sigma^2) = \binom{\mu+\sigma}{\mu} x^\mu (1-x)^{1-\sigma^2}$

2. **Eigenvectors and Eigenvalues**

   If $\boldsymbol{v}$ is an eigenvector of square matrix $\boldsymbol{A}$ with eigenvalue λ, which of the following is true? Boldface lowercase letters like $\boldsymbol{v}$ are column vectors.

   (A) $\boldsymbol{A}\boldsymbol{v} = \boldsymbol{v}^\lambda$

   (B) $\boldsymbol{A}\boldsymbol{v} = \lambda\boldsymbol{v}$

   (C) $\lambda\boldsymbol{A} = \boldsymbol{A}\boldsymbol{v}$

   (D) $\lambda\boldsymbol{A}\boldsymbol{v} = \boldsymbol{v}$

3. **Computing Gradients**

   Which of the following is the gradient of the scalar function $f(\boldsymbol{x}) = (\boldsymbol{a}-\boldsymbol{x})^\top \boldsymbol{B}(\boldsymbol{a}-\boldsymbol{x})$, where $\boldsymbol{B}$ is symmetric? The notation $\boldsymbol{x}^\top$ indicates transpose.

   (A) $\nabla_{\boldsymbol{x}} f = -2\boldsymbol{B}^{-1}(\boldsymbol{a}-\boldsymbol{x})$

   (B) $\nabla_{\boldsymbol{x}} f = -2\boldsymbol{B}(\boldsymbol{a}-\boldsymbol{x})$

   (C) $\nabla_{\boldsymbol{x}} f = 2\boldsymbol{B}^{-1}(\boldsymbol{a}-\boldsymbol{x})$

   (D) $\nabla_{\boldsymbol{x}} f = \boldsymbol{B}^{-1}(\boldsymbol{a}-\boldsymbol{x})$

4. **Properties of Expectations**

Let X and Y be independent random variables with finite expectations. Which of the following statements is false?

(A) $\mathbb{E}[X+Y] = \mathbb{E}[X] + \mathbb{E}[Y]$

(B) $\mathbb{E}[X \cdot Y] = \mathbb{E}[X] \cdot \mathbb{E}[Y]$

(C) $\mathbb{E}[aX + bY] = a^2\mathbb{E}[X] + b^2\mathbb{E}[Y]$

(D) All of these are true.

5. **Basic Graph Theory**

Let G be a bipartite graph with N vertices, where N is even. (A bipartite graph has two groups of vertices and edges can only connect vertices in different groups.) What is the maximum number of edges that G can have?

(A) $N(N-1)/2$

(B) N^2

(C) $N^2/4$

(D) $N(N-1)/4$

6. **Basic Probability**

Which of the following probability identities is false?

(A) $\Pr(A, B) = \Pr(A \mid B)\ \Pr(B)$

(B) $\Pr(A \mid B) = \Pr(B \mid A)\ \Pr(A) / \Pr(B)$

(C) $\Pr(B \mid A) = \Pr(A, B) / \Pr(A)$

(D) All of these are true.

7. **Integral Calculus**

What is the value of the integral

$$\int_{-\infty}^{\infty} f(x)\, \delta(x-a)\, \mathrm{d}x$$

where $\delta(z)$ is the Dirac delta function?

(A) $f(a)$

(B) $f(-a)$

(C) a

(D) $f(0)$

Solutions

1. **The Gaussian/Normal Distribution**
   The correct answer is (A): $\mathcal{N}(x \mid \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$.
   Why is this relevant to a machine learning class? Obviously, you can look up the PDF of a Gaussian distribution on Wikipedia anytime you want, and this class isn't about memorization. However, we'll often take it for granted that you're familiar with this distribution and its properties from, e.g., Stat 110. We'll use the Gaussian distribution, for example, when we talk about regression errors and when we study mixture models for clustering.

2. **Eigenvectors and Eigenvalues**
   The correct answer is (B): $\boldsymbol{A}\boldsymbol{v} = \lambda \boldsymbol{v}$.
   Why is this relevant to a machine learning class? It's very important to have a good intuition about eigenvectors, because we often use them to understand the underlying low-dimensional structure of our problems. For example, principal component analysis (PCA) relies heavily on eigendecomposition.

3. **Computing Gradients**
   The correct answer is (B): $\nabla_{\boldsymbol{x}} f = -2\boldsymbol{B}(\boldsymbol{a} - \boldsymbol{x})$.
   Why is this relevant to a machine learning class? A lot of machine learning is just about computing gradients. We come up with a loss function in terms of parameters and data, and then want to minimize it. We do this by computing the gradient of the loss in terms of the parameters and either analytically or numerically finding a solution. This comes up in regression, neural networks, K-Means clustering, and many more settings. Sometimes machine learning feels like it is nothing more than just computing gradients!

4. **Properties of Expectations**
   The correct (that is, false statement) answer is (C): $\mathbb{E}[aX + bY] = a^2\mathbb{E}[X] + b^2\mathbb{E}[Y]$. Expectation is linear, so the true version would be $\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y]$.
   Why is this relevant to a machine learning class? A lot of machine learning is about reasoning under uncertainty, and to do that we need to talk about expectations of various things. We might talk about expected prediction loss, or about the expected value of some feature, or about the expected utility of taking some action – these are all things that will come up this semester.

5. **Basic Graph Theory**
   The correct answer is (C): $N^2/4$. The graph has the most edges when the groups have the same size and are fully connected between them.
   Why is this relevant to a machine learning class? Graph theory comes up quite often in machine learning, particularly in the context of probabilistic modeling. In fact, there is an entire subfield dedicated to connecting graph theory and probability theory, called *probabilistic graphical models*. We'll discuss this when we look at hidden Markov models. Graphs come up also when examining neural networks, and in many other machine learning subareas, such as nonlinear dimensionality reduction.

6. **Basic Probability**
   The correct answer is (D): they are all true. These are all applications of the product rule of

probability.
Why is this relevant to a machine learning class? The field of machine learning depends on probability and statistics so heavily that it is often call *statistical* machine learning. Probability and information theory figure heavily into almost everything we'll do in this course.

7. **Integral Calculus**
   The correct answer is (A): $f(a)$. The Dirac delta is a generalized function essentially defined by

$$\int_{-\infty}^{\infty} f(x)\,\delta(x)\,\mathrm{d}x = f(0).$$

   Why is this relevant to a machine learning class? Integral calculus and concepts like delta functions come up a lot because we spend a lot of time thinking about probabilities. Delta functions are a way to write down something that is deterministic as kind of degenerate distribution. These ideas also come up when we talk about convolutional neural networks.