Question 1

The amount of traffic in winter in Providence depends on whether it snows or not, and it snows with probability $\frac{1}{3}$. With no snow, the time for a taxi to drive from Brown to the airport is exponentially distributed with a mean of 15 minutes. With snow, the travel time is exponentially distributed with a mean of 25 minutes. Given that it takes you 23 minutes to reach the airport, what is the probability that it is snowing?

Answer 1
Question 2

In a 90-minute soccer match, let $G = 1$ if at least one goal is scored, and $G = 0$ if no goals are scored (the game ends in a zero-zero tie). Define the continuous random variable $X$ to be the time that the first goal is scored. In 90% of games, one or more goals are scored ($G = 1$), and $X$ has the following probability density function for some constant $c > 0$:  

$$f_{X|G}(x|G = 1) = \begin{cases}  
    cx & \text{if } 0 \leq x \leq 90, \\
    0 & \text{otherwise.}
\end{cases}$$

In the 10% of games where no goals are scored ($G = 0$), we let $X = 90$.

(a) Find the value of $c$ for which $f_{X|G}(x|G = 1)$ is a valid probability density function. Determine a functional form for the corresponding cumulative distribution function.

(b) Determine the mean and variance of $X$ given that $G = 1$. Simplify your answer.

(c) Determine the mean and variance of $X$. Simplify your answer.

(d) Find the time $x$ (in minutes) at which there is an 80% probability that at least one goal has been scored.

(e) Suppose that after 45 minutes (at halftime), no goals have been scored. Given this knowledge, what is the probability that no goals are scored in the rest of the 90-minute match?

Answer 2
Question 3

An online computer-science class has an assignment to write a chess-playing program that estimates the optimal move given the state of an ongoing game. Suppose that the runtimes of student programs follow a normal distribution with mean $\mu = 15$ seconds, and standard deviation $\sigma = 2.5$ seconds. *Hint:* The Matlab commands `normcdf` and `norminv` may be useful for this question.

(a) What is the probability that a random program has a runtime greater than 19 seconds?

(b) What is the probability that a random program has a runtime between 12 and 18 seconds?

(c) The TAs want to help the students complete their work faster. What would they have to lower the average runtime to so that only 1.0% of students have runtimes over 15 seconds? Assume the standard deviation remains fixed at $\sigma = 2.5$ seconds.

(d) The technical support staff installs a new online server for running student experiments, and now the runtime of each program is exactly $1/5$ of what it used to be. Given this new server (but not extra TA help as in part (c)), what are the mean and standard deviation of the new runtime distribution?

Answer 3
Question 4

For a given day \( i \), we let \( Y_i = 1 \) if the ground-level ozone concentration near some city (Houston, in our data) is at a dangerously high level. This is called an “ozone day.” We let \( Y_i = 0 \) if the ozone concentration is low enough to be considered safe.

We want to predict \( Y_i \) from more easily measured “features” describing atmospheric pollutant levels and meteorological conditions (temperature, humidity, wind speed, etc.). There are a total of \( M = 72 \) of these features collected each day, which we denote by \( X_{ij} \). Each feature \( X_{ij} \in \mathbb{R} \) is a real number, and we will thus use a Gaussian distribution to model these continuous random variables.

We will build a “naive Bayes” classifier, which labels observation \( i \) as an ozone day if \( P(Y_i = 1|X_i) > P(Y_i = 0|X_i) \), and as a non-ozone day otherwise. Using Bayes rule, this classifier is equivalent to one that chooses \( Y_i = 1 \) if and only if

\[
\frac{p_Y(1)f_{X|Y}(x_i|y_i = 1)}{f_X(x_i)} > \frac{p_Y(0)f_{X|Y}(x_i|y_i = 0)}{f_X(x_i)},
\]

or equivalently

\[
\log p_Y(1) + \log f_{X|Y}(x_i|y_i = 1) > \log p_Y(0) + \log f_{X|Y}(x_i|y_i = 0). \quad (1)
\]

In this equation, \( p_Y(y_i) \) is the probability mass function that defines the prior probability of ozone and non-ozone days. The conditional probability density function \( f_{X|Y}(x_i|y_i) \) describes the distribution of the \( M = 72 \) environmental features, which we assume depends on the type of day. We make two simplifying assumptions about these densities: the features \( X_{ij} \) are conditionally independent given \( Y_i \), and their distributions are Gaussian. Thus:

\[
f_{X|Y}(x_i|y_i = 1) = \prod_{j=1}^{M} \frac{1}{\sqrt{2\pi\sigma_{1j}^2}} \exp \left\{ -\frac{(x_{ij} - \mu_{1j})^2}{2\sigma_{1j}^2} \right\}, \quad (2)
\]

\[
f_{X|Y}(x_i|y_i = 0) = \prod_{j=1}^{M} \frac{1}{\sqrt{2\pi\sigma_{0j}^2}} \exp \left\{ -\frac{(x_{ij} - \mu_{0j})^2}{2\sigma_{0j}^2} \right\}. \quad (3)
\]

Given \( Y_i = 1 \), \( X_{ij} \) is Gaussian with mean \( \mu_{1j} \) and variance \( \sigma_{1j}^2 \). Given \( Y_i = 0 \), \( X_{ij} \) is Gaussian with mean \( \mu_{0j} \) and variance \( \sigma_{0j}^2 \). There are a total of \( 2M \) mean parameters and \( 2M \) variance parameters, since every feature \( X_{ij} \) has a distinct distribution for each class \( Y_i \).

(a) Derive equations for \( \log f_{X|Y}(x_i|y_i = 1) \) and \( \log f_{X|Y}(x_i|y_i = 0) \), the logarithms of the conditional probability density functions in Equations (2,3). For numerical robustness, simplify your answer so that it does not involve the exponential function.

Because ozone days are relatively rare, a classifier that always predicts \( Y_i = 0 \) would be correct over 95% of the time, but would obviously not be practically useful for reducing ozone hazard. To evaluate our classifiers, we will thus separately compute the numbers of false alarms (predictions of ozone days when in reality \( Y_i = 0 \)) and missed detections (predictions of non-ozone days when in reality \( Y_i = 1 \)). We are willing to allow some false alarms as long as there are very few missed detections.

For all parts below, assume that the mean parameters \( \mu_{1j} \), \( \mu_{0j} \) are set to match the mean of the empirical distribution of the training data. The demo code computes these means.
(b) Start by assuming the classes are equally probable \((p_Y(1) = p_Y(0) = 1/2)\), and have unit variance \((\sigma_{1j}^2 = \sigma_{0j}^2 = 1)\). Write code to compute the log conditional densities from part (a). Then using Equation (1), classify each test example. Report your classification accuracy, and the numbers of false alarms and missed detections.

Hint: Your classifier should have fewer than 10 missed detections.

(c) Rather than assuming features have variance one, set the variance parameters \(\sigma_{1j}^2, \sigma_{0j}^2\) equal to the variance of the empirical distribution of the training data. Classify each test example using Equation (1) with these variance estimates. Report your classification accuracy, and the numbers of false alarms and missed detections.

(d) Rather than assuming the classes are equally probable, estimate \(p_Y(1)\) as the fraction of training examples that are ozone days. Classify each test example using Equation (1) with this informative class prior, and the variances from part (c). Report your classification accuracy, and the numbers of false alarms and missed detections.

(e) Suppose we assume unit variance but use the \(p_Y(1)\) estimates of part (d). How do you think this classifier would differ from that of (a)? Do you expect more or less false positives and more or less true negatives, and why? Which estimator would be better for this particular use case?

**Answer 4**