Problem 1. (5 points) Let \((\Omega, \mathcal{A}, P)\) be a probability space and \(A \subseteq \Omega\) and \(B \subseteq \Omega\) any two subsets of \(\Omega\). Prove the following expression or provide a counterexample if it does not hold
\[ P(A) = P(A|B) + P(A|B^c), \]
where \(A^c\) is the complement of \(A\).

Problem 2. (15 points) Let \(X\) be a random variable on \(X = \{a, b, c\}\) with the probability mass function \(p(x)\). Let \(p(a) = 0.1\), \(p(b) = 0.2\), and \(p(c) = 0.7\) and some function \(f(x)\) be
\[
f(x) = \begin{cases} 
10 & x = a \\
5 & x = b \\
\frac{10}{7} & x = c
\end{cases}
\]

a) (5 points) What is \(E[f(X)]\)?
b) (5 points) What is \(E[1/p(X)]\)?
c) (5 points) For an arbitrary finite set \(\mathcal{X}\) with \(n\) elements and arbitrary \(p(x)\) on \(\mathcal{X}\), what is \(E[1/p(X)]\)?

Problem 3. (10 points) Let \(X\) and \(Y\) be random variables. Prove or disprove the following formula
\[ V[X + Y] = V[X] + V[Y] + 2\text{Cov}[X,Y], \]
where \(V[X]\) is the variance of \(X\) and \(\text{Cov}[X,Y]\) is the covariance between \(X\) and \(Y\).

Problem 4. (20 points) Suppose that the number of accidents occurring daily in a certain plant has a Poisson distribution with an unknown mean \(\lambda\). Based on previous experience in similar industrial plants, suppose that our initial feelings about the possible value of \(\lambda\) can be expressed by an exponential distribution with parameter \(\theta = \frac{1}{2}\) is, the prior density is
\[ p(\lambda) = \theta e^{-\theta \lambda} \]
where \(\lambda \in (0, \infty)\). If there are 99 accidents over the next 11 days, determine

a) (5 points) the maximum likelihood estimate of \(\lambda\)
b) (5 points) the maximum a posteriori estimate of \(\lambda\)
c) (10 points) the Bayes estimate of \(\lambda\).
Problem 5. (10 points) Let $D = \{x_i\}_{i=1}^n$ be an i.i.d. sample from
\[ p(x) = \begin{cases} e^{-(x-\theta_0)} & x \geq \theta_0 \\ 0 & \text{otherwise} \end{cases} \]
Determine $\theta_{ML}$ - the maximum likelihood estimate of $\theta_0$.

Problem 6. (25 points) Let $D = \{x_i\}_{i=1}^n$, where $x_i \in \mathbb{R}$, be a data set drawn independently from a Gumbel distribution
\[ p(x) = \frac{1}{\beta} e^{-\frac{x-\alpha}{\beta}} e^{-e^{-\frac{x-\alpha}{\beta}}}, \]
where $\alpha \in \mathbb{R}$ is the location parameter and $\beta > 0$ is the scale parameter.

a) (10 points) Derive an algorithm for estimating $\alpha$ and $\beta$.

b) (10 points) Implement the algorithm derived above and evaluate it on data sets of different sizes. First, find or develop a random number generator that creates a data set with $n \in \{100, 1000, 10000\}$ values using some fixed $\alpha$ and $\beta$. Then make at least 10 data sets for each $n$ and estimate the parameters. For each $n$, report the mean and standard deviation on the estimated $\alpha$ and $\beta$. If $n = 10000$ is too large for your computing resources, skip it.

c) (5 points) The problem above will require you to implement an iterative estimation procedure. You will need to decide on how to initialize the parameters, how to terminate the estimation process and what the maximum number of iterations should be. Usually, some experimentation will be necessary before you run the experiments in part (b) above. Summarize what you did in a short paragraph, no more than two paragraphs.

NB: There are several versions and naming conventions for the Gumbel distribution in the literature.

Problem 7. (30 points) Let $D = \{x_i\}_{i=1}^n$, where $x_i \in \mathbb{R}$, be a data set drawn independently from the mixture of two distributions
\[ p(x) = w_1 p_1(x) + w_2 p_2(x), \]
where
\[ p_1(x) = \frac{1}{\beta} e^{-\frac{x-\alpha}{\beta}} e^{-e^{-\frac{x-\alpha}{\beta}}}, \]
is a Gumbel distribution with parameters $\alpha$ and $\beta$,
\[ p_2(x) = \frac{1}{\sqrt{2\pi}\sigma^2} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \]
is a Gaussian distribution with parameters $\mu$ and $\sigma$, and $(w_1, w_2)$ are positive constants such that $w_1 + w_2 = 1$.

a) (15 points) Derive an EM algorithm for estimating $w_1$, $w_2$, $\alpha$, $\beta$, $\mu$ and $\sigma$.

b) (15 points) Implement the algorithm derived above and evaluate it on data sets of different sizes. Use similar experimentation as in Problem 3.

You can recycle derivations and code from Problem 6. You can also use any result and derivation from the lecture notes posted on the class web site.
Problem 8. (20 points) Understanding the curse of dimensionality. Consider the following experiment: generate \( n \) data points with dimensionality \( k \). Let each data point be generated using a uniform random number generator with values between 0 and 1. Now, for a given \( k \), calculate

\[
r(k) = \log_{10} \frac{d_{\text{max}}(k) - d_{\text{min}}(k)}{d_{\text{min}}(k)}
\]

where \( d_{\text{max}}(k) \) is the maximum distance between any pair of points and \( d_{\text{min}}(k) \) is minimum distance between any pair of points (you cannot use identical points to obtain the minimum distance of 0). Let \( k \) take each value from \( \{1, 2, \ldots, 99, 100\} \). Repeat each experiment multiple times to get stable values by averaging the quantities over multiple runs for each \( k \).

a) (15 points) Plot \( r(k) \) as a function of \( k \) for two different values of \( n \); \( n \in \{100, 1000\} \). Label and scale each axis properly to be able to make comparisons over different \( n \)'s. Embed your final picture(s) in the file you are submitting for this assignment.

b) (5 points) Discuss your observations and also compare the results to your expectations before you carried out the experiment.
Homework Directions and Policies

Submit a single package containing all answers, results and code. Your submission package should be compressed and named firstnamelastname.zip (e.g., predragradivojac.zip). In your package there should be a single pdf file named main.pdf that will contain answers to all questions, all figures, and all relevant results. Your solutions and answers must be typed\footnote{We recommend Latex; in particular, TexShop-MacTeX combination for a Mac and TeXnicCenter-MiKTex combination on Windows. An easy way to start with Latex is to use the freely available Lyx. You can also use Microsoft Word or other programs that can display formulas professionally.} and make sure that you type your name and Northeastern username (email) at the beginning of the file. The rest of the package should contain all code that you used. The code should be properly organized in folders and subfolders, one for each question or problem. All code, if applicable, should be turned in when you submit your assignment as it may be necessary to demo your programs to the teaching assistants. Use Matlab, Python, or C/C++.

Unless there are legitimate circumstances, late assignments will be accepted up to 5 days after the due date and graded using the following rules:

- on time: your score $\times 1$
- 1 day late: your score $\times 0.9$
- 2 days late: your score $\times 0.7$
- 3 days late: your score $\times 0.5$
- 4 days late: your score $\times 0.3$
- 5 days late: your score $\times 0.1$

For example, this means that if you submit 3 days late and get 80 points for your answers, your total number of points will be $80 \times 0.5 = 40$ points.

All assignments are individual, except when collaboration is explicitly allowed. All the sources used for problem solution must be acknowledged; e.g., web sites, books, research papers, personal communication with people, etc. Academic honesty is taken seriously! For detailed information see Office of Student Conduct and Conflict Resolution.