1. The following table shows the quiz 1 and quiz 2 grades obtained for students in a database course.

x	y
Quiz 1	Quiz 2
72	84
50	63
81	77
74	78
94	90
86	75
59	49
83	79
65	77
33	52
88	74
81	90

Table 1: Example dataset

For different choices of the feature function ϕ (such as log, exponentiation etc), construct decision trees to predict a students' quiz 2 marks based on the students' quiz 1 marks in the course. Does the choice of ϕ matter?

2. Suppose that we have three text documents d_1, d_2, and d_3. Document d_1 contains 30 nouns, 40 verbs, and 30 adjectives document d_2 contains 1 nouns, 1 verbs, and 0 adjectives, and box d_3 contains 3 nouns, 3 verbs, and 4 adjectives. If a document is chosen at random with probabilities p(d_1) = 0.2, p(d_2) = 0.2, p(d_3) = 0.6, and a word is removed from the document (with equal probability of selecting a word of any of the parts of speech in the document), then what is the probability of selecting a noun? If we observe that the selected word is in fact a verb, what is the probability that it came from d_3?

   (This is problem number 1.3 from Bishop's book slightly reworded... It talks about choosing fruits from boxes).

3. For this exercise, refer to the discussion in class on information content for a probability distribution (Section 9.1.2 of last year class notes). Prove that

   (a) For a discrete random variable (with countable domain), the information is maximum for the uniform distribution.

(b) For Continuous random variable (with finite mean and finite variance), the information is maximum for the Gaussian Distribution.

4. **Part of speech(pos) example:** Problem Statement:-
A set of 'n' words, each of a particular part of speech(noun/verb/etc) is picked. Probability that a word is of part of speech type 'k' is p_k. Assuming the picking of words is done independently, find probability that the set contains a 'noun' given that it contains a 'verb'.

Solution
Let A_k be the probability that the set contains pos type 'k'.
$Pr(A_k) = 1 - (1 - p_k)^n$
where $(1 - p_k)^n$ is that all 'n' words are not of pos of type 'k'.

$Pr(A_{noun}/A_{verb}) = \frac{Pr(A_{noun} \bigcap A_{verb})}{Pr(A_{verb})}$

$Pr(A_{k1} \bigcap A_{k2}) = 1 - (1 - p_{k1})^n - (1 - p_{k2})^n + (1 - p_{k1} - p_{k2})^n$

$Pr(A_{noun}/A_{verb}) = \frac{1-(1-p_{noun})^n-(1-p_{verb})^n+(1-p_{noun}-p_{verb})^n}{1-(1-p_{verb})^n}$

5. X and Y are independent continuous random variables with same density functions.

$$p(x) = \begin{cases} e^{-x} & \text{if } x > 0; \\ 0 & \text{otherwise.} \end{cases}$$

Find density $\frac{X}{Y}$.
Note:- They are indepedent.

Solution

$F_{\frac{X}{Y}}(a) = Pr(\frac{X}{Y} <= a)$
$= \int_0^\infty \int_0^{ya} p(x, y) dx dy$
$= \int_0^\infty \int_0^{ya} e^{-x} e^{-y} dx dy$
$= 1 - \frac{1}{a+1}$
$= \frac{a}{a+1}$

$f_{\frac{X}{Y}}(a) =$ derivative of $F_{\frac{X}{Y}}(a)$ w.r.t a
$= \frac{1}{(a+1)^2} > 0$

6. For the problem of maximizing likelihood given a random sample $\{X_1, X_2, \ldots, X_n\}$ generated by a multivariate bernoulli $\sim (\mu_1, \mu_2, \ldots, \mu_k)$ (which we have been discussing since some time in the class)

$$
\begin{aligned}
\widehat{\mu} = \quad & \underset{\mu_1,\mu_2,\ldots,\mu_k}{\text{argmax}} && \sum_{j=1}^{k} n_j \log(\mu_j) \\
& \textit{such that} && \sum_{j=1}^{k} \mu_j = 1 \\
& \textit{and} && \mu_j \in [0,1]\ \forall j = 1,\ldots,k
\end{aligned}
$$

where n_j is the number of instances in the random sample that assume the value v_j.

Applying the necessary conditions for optimality from Section 4.4 of `http://www.cse.iitb.ac.in/~cs725/notes/classNotes/BasicsOfConvexOptimization.pdf` to this optimization problem, compute all the possible values of $\widehat{\mu}$.

7. Consider the half space defined by $\mathcal{H} = \left\{\mathbf{x} \in \Re^n | \mathbf{a}^T\mathbf{x} + \alpha \geq 0\right\}$ where $a \in \Re^n$ and $\alpha \in \Re$ are given. Formulate and solve the optimization problem for finding the point $\mathbf{x}$ in $\mathcal{H}$ that has the smallest Euclidean norm.

8. You should redo all the problems in Section 4.1 from `http://www.cse.iitb.ac.in/~cs725/notes/classNotes/BasicsOfConvexOptimization.pdf`.

9. Extend the proof for problem 2 in Quiz 1 to the case of multiple classes.

10. Solve Exercises 2.1, 2.3, 2.4, 2.5, 2.9, 2.11, 2.14, 2.17, 2.60, 2.40, 2.47, 2.56, 8.3, 8.4, 4.9. Solutions can be found at `http://research.microsoft.com/en-us/um/people/cmbishop/prml/prml-web-sol-2009-09-08.pdf`.

11. Prove expressions for the posterior of mean $P(\mu \| x_1, x_2, \ldots x_m)$ of the Gaussian distribution under Gaussian prior for means (with fixed known variance) as in section 10.2 on bottom of page 50 of `http://www.cse.iitb.ac.in/~cs725/notes/classNotes/lecturenote_2010.pdf`. Also prove the expressions for μ_m and σ_m^2.

12. Solve problem 5 from `http://www.stanford.edu/class/cs229/materials/ps3.pdf`. Solve all parts a, b and c. Solution to this can be found at `http://www.stanford.edu/class/cs229/restricted/ps3-sol.pdf`.

13. Solve problem 4 from `http://www.stanford.edu/class/cs229/restricted/ps3-sol.pdf`. Solve all parts a, b, c. Solution to this can be found at `http://www.stanford.edu/class/cs229/restricted/ps1.pdf`.