CPSC-340 Machine Learning and Data Mining 2008

Homework # 3

Due Monday, Oct 13 1pm.

NAME:______________________________________

Signature:_____________________________________

STD. NUM: ___________________________________

General guidelines for homeworks:

You are encouraged to discuss the problems with others in the class, but all write-ups are to be done on your own.

Homework grades will be based not only on getting the "correct answer," but also on good writing style and clear presentation of your solution. It is your responsibility to make sure that the graders can easily follow your line of reasoning.

Try every problem. Even if you can't solve the problem, you will receive partial credit for explaining why you got stuck on a promising line of attack. More importantly, you will get valuable feedback that will help you learn the material.

Please acknowledge the people with whom you discussed the problems and what sources you used to help you solve the problem (e.g. books from the library). This won't affect your grade but is important as academic honesty.

When dealing with python exercises, please attach a printout with all your code and show your results clearly.

1. **(Probability revision)**

   (i) Let $P(HIV = 1) = 1/500$ be the probability of contracting HIV by having unprotected sex. If one has unprotected sex twice, what is the probability of contracting HIV? What if we have unprotected sex 500 times?

(ii) Let $A \in \mathbb{R}^{k \times n}$, $b \in \mathbb{R}^k$ be given matrices, and $X \in R^n$ be a random variable with mean $\mathbb{E}(X) = \mu_x \in R^n$ and covariance $cov(X) = \Sigma_X \in \mathbb{R}^{n \times n}$. We define a new random variable

$$Y = AX + b$$

If $X \sim N(\mu_x, \Sigma_x)$, show that $Y \sim N(\mu_y, \Sigma_y)$ by deriving expressions for μ_y and Σ_y in terms of the sufficient statistics (μ_x, Σ_x) of X and the parameters of the linear transformation (A, b).

(iii) Let $x_1 \sim N(\mu_1, \sigma_1^2)$ and $x_2 \sim N(\mu_2, \sigma_2^2)$ be two **independent** random variables, derive an expression for $p(x_1, x_2)$. Explain the zeros in the covariance matrix of the joint distribution of x_1 and x_2.

(iv) A random variable X with a uniform distribution between 0 to 1 is written as $X \sim \mathcal{U}_{[0,1]}(x)$. Draw pictures of the pdf and cdf of this random variable in the one-dimensional case. Draw a picture of the pdf of X when it is two-dimensional.

(v) Let $X \sim N(\mu, \Sigma)$ and let the eigenvalue decomposition of the covariance be $\Sigma = U\Lambda U^T$. Prove that $X \sim \mu + U\Lambda^{1/2}N(0, I)$. Provide a geometric interpretation of the effect of U and Λ in 2D. How does this relate to PCA?

(vi) For the set indicator variable $\mathbb{I}_{\mathbb{A}}(\omega)$, complete the following:

$$\mathbb{E}[\mathbb{I}_{\mathbb{A}}(\omega)] =$$

2. **(PCA)** This question is for extra marks (extra 5% of midterm1). As explained in class in the context of digit images (the images of 2's), you'll have to implement a PCA projection of, say 10, images to a single 2D display. Use any images you might like to display on the PCA layout.