Sample questions for Test 1 for CSI 431

September 23, 2020

1. **Statistics:** Let X and Y be two random variables. Consider a random sample of size $n = 6$ from these two variables $X = (1, 4, 6, 2, 2, 3)$ and $Y = (5, 8, 3, 12, 14, 18)$.
   (a) Find the mean and median of X.
   (b) Find the variance of Y.
   (c) Estimate the probability of observing a value of at most 13 for Y?
   (d) Find the 2-dimensional mean $\hat{\mu}$ and the covariance matrix $\hat{\Sigma}$ for these two variables.
   (e) What is the correlation between X and Y?

2. **CDFs and Means** Consider two univariate discrete random variables A and B with means $\mu_A < \mu_B$, where both variables can take on values in $\{1, 2, 3\}$. Is it true that the CDF of A always dominates that of B. If yes, prove it. If not, give example PMFs for A and B in which the statement is incorrect.

3. **Outliers:** Let X be a random variable with a sample of values $A = (1, 7, 2, 7, 3, 4)$. What is the minimum value of an additional sample you need to add in order to make the the new mean at least twice the original mean? Would the median also double in value? Show your computations.

4. **Traces:** Prove that $tr(AB) = tr(BA)$ if A and B are square matrices of dimension $n \times n$ and $tr()$ is the trace operator defined as the sum of the diagonal elements of a matrix.

5. **Rank:** Give an example of a 3x3 matrix with a column rank of 2

6. **Multiplication:** If possible, compute the inner and outer products of the column vectors $x = (1, 0, 3, 4, 5)^T$ and $y = (1, 0, 2)^T$

7. **Eigenvectors:** Is vector $u = \begin{bmatrix} 1/\sqrt{2} \\ 0 \\ -1/\sqrt{2} \end{bmatrix}$ an eigenvector of matrix $A = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}$? Prove or disprove. If it is, show the corresponding eigenvalue?

8. **Eigenvalue decomposition:** Given the covariance matrix $A = \begin{bmatrix} 5 & 4 \\ 4 & 5 \end{bmatrix}$, answer the following questions:
(a) Compute the eigenvalues of A by solving the equation $det(A - \lambda I) = 0$.
(b) Find the corresponding eigenvectors by solving $Au_i = \lambda_i u_i$.

9. **PCA:** Given the covariance matrix $A = \begin{bmatrix} 5 & 4 \\ 4 & 5 \end{bmatrix}$, and reusing your computations from the Eigenvalue decomposition questions above, answer the following questions:
(a) What is the % of retained variance if keeping only the first principal component?
(b) Come up with a dataset of 4 points which could have resulted in the covariance matrix above.
(c) Project your dataset from the previous part onto the principal component (only the first) and show the new coordinate of the points.

10. **SVD:** Using the eigendecomposition of the covariance matrix $A = \begin{bmatrix} 5 & 4 \\ 4 & 5 \end{bmatrix}$ from the previous two problems, what would be the first two right singular vectors of a dataset which resulted in this covariance matrix? What would be their corresponding singular values? *Hint: use the relationship between PCA and SVD we discussed in class.*

11. **Kernel methods:** Given the following data matrix $A = \begin{bmatrix} 5 & 1 \\ 3 & 5 \\ 2 & 5 \\ 0 & 1 \end{bmatrix}$, and the RBF kernel: $K(x, y) = e^{-||x-y||^2}$ answer the following questions:
(a) Compute the kernel matrix K
(b) Find the farthest pair of points and closest pair of points in kernel space in terms of Euclidean distance, i.e. $||\phi(x) - \phi(y)||$. Are these pairs the same in the original input space?

12. **Generic question:** Pick a slide where we worked out an example in class (only topics for the test), change the example (say input data, or split in DT, or some other aspect of the example), solve the new example. To post a solution post it as Question11-SlideXXX-TopicYYY, where XXX is the slide number and YYY is the topic (slide deck). Post both the new example and your solution.