Department of Electrical Engineering, IIT Delhi

EEL709 Pattern Recognition: Minor II Examination

(Closed book/Closed Notes) Time: 1 hour   Maximum Marks: 20

"Thou shalt not covet thy neighbour's answers"

Useful Formulae and Results:

The KLT: $\mathbf{r_i} = \mathbf{U}^T\ \mathbf{p_i}$. Here $\mathbf{p_i}$ are Type-I normalised $k \times 1$ patterns (n of which can be stacked to get the Type-I normalised pattern matrix $\mathbf{P}$). $\mathbf{U}$ is a matrix of eigenvectors of the covariance matrix $\mathbf{A} = \frac{1}{n}\mathbf{P}\mathbf{P}^T$. There are k eigenvectors $\mathbf{u_i}$ corresponding to eigenvalues λ_i.

Useful Result 1: Eigenvectors of a symmetric matrix correspoding to non-repeated eigenvalues, are orthonormal.

Useful Result 2: Diagonalisation of a square matrix $\mathbf{B}$: $\mathbf{B} = \mathbf{U}\Lambda\mathbf{U}^{-1}$.

The SVD: $\mathbf{P} = \mathbf{U}\ \Sigma\ \mathbf{V}^T$, where this $\mathbf{U}$ is the $k \times k$ matrix of orthonormal basis vectors, Σ is a $k \times n$ matrix having singular values σ_i along the main diagonal (the other values are all zero), and $\mathbf{V}$ is an $n \times n$ matrix of the eigenvectors of $\mathbf{A'} = \mathbf{P}^T\mathbf{P}$. These eigenvectors $\mathbf{v_i}$ correspond to eigenvalues λ_i, and we define $\sigma_i = \sqrt{\lambda_i}$, and $\mathbf{u_i} = \frac{1}{\sigma_i}\mathbf{P}\mathbf{v_i}$.

Useful Result 3: Gram-Schmidt Orthogonalisation: $\mathbf{u_k} = \mathbf{v_k} - \sum_{j=1}^{k-1} \frac{\langle \mathbf{u_j}, \mathbf{v_k} \rangle}{\langle \mathbf{u_j}, \mathbf{u_j} \rangle} \mathbf{u_j}$, where $\mathbf{u_j}$ are the new orthogonal basis vectors corresponding to the given set of k basis vectors $\mathbf{v_j}$, $j = 1, \dots k$, and all other symbols have their usual meaning.

---

1. **कभी आयी खुशी, कभी Eigen...**

   (a) **सफ़ेदी की छमकार, ज्यादा सफ़ेद - The Whitening Transformation:** This is given by $\tilde{\mathbf{r}}_\mathbf{i} = \Lambda^{-1/2}\ \mathbf{U}^T\ \mathbf{p_i}$. Here, we deal with $k \times 1$ Type-I normalized pattern vectors (*i.e.*, normalised with respect to the mean) $\mathbf{p_i}$, $0 \leq i \leq n-1$. Λ is a diagonal matrix of their eigenvalues, and $\mathbf{U}$ is their corresponding eigenvector matrix. What is the covariance matrix of the transformed patterns $\tilde{\mathbf{r}}_\mathbf{i}$? *(2 marks)*

   (b) **Orthopaedic Orthonormality: Bone-Breaking Normal Work** Show that for non-repeated eigenvalues of a symmetric matrix, the eigenvectors will be orthonormal. Consider the case of repeated eigenvalues of a symmetric matrix $\mathbf{A}$. Specifically use the Gram-Schmidt orthogonalisation process to show that it is still possible to obtain a set of orthonormal vectors. *(5 marks)*

   (c) **Projection Interjection, Nothing to Lose?** What is the physical significance of projecting a new vector onto an eigenspace? Explain using mathematical expressions, what the above implies, for both the KL-Transform as well as the SVD. Now, explain how the KL-Transform and the SVD respectively perform lossy compression. *(2 + 2 marks)*

2. The *logistic regression* model for binary classification applies the following mapping for a given input feature vector $\mathbf{x}$:

$$p(C_1|\mathbf{x}) = y(\mathbf{x}) = \sigma(\mathbf{w}^{\mathrm{T}}\mathbf{x});$$

   where $\mathbf{w}$ is a vector of parameters (including an intercept w_0, with the convention that $x_0 = 1$), σ is the logistic sigmoid function, and C_1 is one of the two class labels.

Write down the likelihood of a given data set $\{(\mathbf{x}_1, t_1); (\mathbf{x}_2, t_2); ...; (\mathbf{x}_N, t_N)\}$, where the output $t_n \in \{0, 1\}$, corresponding to the classes labeled C_0 and C_1 respectively. Show that for a linearly separable data set, the maximum likelihood solution is given by a vector $\mathbf{w}$ whose decision boundary corresponds to a separating hyperplane, and where the magnitude of $\mathbf{w}$ is infinity. *(4 marks)*

3. Consider a binary classification problem where the input feature space is 2-dimensional. The coordinates of the positive examples are $(1, 1)$ and $(-1, -1)$. The coordinates of the negative examples are $(1, -1)$ and $(-1, 1)$.

   (a) Is the data separable in the original space? Consider a feature transformation defined by $\boldsymbol{\phi}(\mathbf{x}) = [1, x_1, x_2, x_1x_2]$, where x_1 and x_2 are the two elements of the input vector $\mathbf{x}$. In this transformed space, give the weight vector $\mathbf{w}$ corresponding to the maximum-margin separating hyperplane. (You need not show any derivation for this.) *(2 marks)*

   (b) Now add a fifth data point to your set, such that the data becomes non-separable even in the transformed space. Draw a plot showing the original 4 points plus your new point in the original 2-dimensional feature space. Justify why your new point makes the data non-separable even in the $\boldsymbol{\phi}(\mathbf{x})$ space. *(1.5 marks)*

   (c) Write down the *kernel function* corresponding to $\boldsymbol{\phi}(\mathbf{x})$. Is this a Mercer kernel? *(1.5 marks)*