Lecture 25

Name: Harsh Sanjay Roniyar

Roll Number: 22B3942

Principal Component Analysis

Objective:

- Find a way to select top $d < D$ orthogonal directions that explain the maximum possible variance of the data

Outline:

- Mean-center the data
- Loop
  - Find direction of maximum variance (that is orthogonal to all previously found directions)
- Until desired percent of variance is captured or number of dimensions

Algorithm

- **Mean centering:** $z := x - \mu_x$
- **Covariance:** $NC = Z^T Z$
- **Eigen decomposition:** $C = U\Lambda U^T$; $\lambda_j u_j = Cu_j$
- **Selection:** $C_d = U\Lambda_d U^T = U_d \Lambda_d U_d^T$, $d < D$
  - Drop dimensions: Set lower $D - d$ eigenvalues to zero
- **Projection:** $Y = ZU_d$
- **Reconstruction:** $Z_d = YU_d^T = ZU_d U_d^T$
- **Reconstruction error:** $||Z - Z_d||_2^2 \propto ||\Lambda - \Lambda_d||$

Kernel PCA

- Introduces nonlinearity by mapping data to a feature space
- Avoids calculating features directly by using the kernel trick

t-distributed Stochastic Node Embeddings (T-SNE)

Auto-encoder method

Linear Support Vector Regression: Points outside the margin are the support vectors for hard-regression, contrary to classification support vectors only on the boundaries.

Decision Tree: Pruning,