CS 383 – Machine Learning

Notation

Machine Learning Notation

- C – Number of classes
- k – Number of clusters.
- C_i – The set of data with class or cluster i
- $|C_i|$ – Cardinality (set size) of set C_i
- D – Dimensionality of data vector (number of features)
- R – Number of outputs
- N – Number of data cases (observations, instances)
- X – Complete dataset observations.
- Y – Complete dataset outputs.
- X_i – The i^{th} observation in dataset X, $i \in \{1, \dots, N\}$
- Y_i – The i^{th} observation's output, $i \in \{1, \dots, N\}$
- $X_{:,j}$ – The j^{th} feature of all the data in the dataset X, $j \in \{1, \dots, D\}$
- $Y_{:,j}$ – The j^{th} output for all observations.
- $X_{i,j}$ – The j^{th} feature of observation $i, i \in \{1, \dots, N\}, j \in \{1, \dots, D\}$.
- $Y_{i,j}$ – The j^{th} output for observation $i \in \{1, \dots, N\}$
- x – A single sample observation.
- x_i – The j^{th} feature of observation x.
- y – A single observation's output.
- y_i – The i^{th} output for an observation.
- θ – parameter vector
- θ_s – s^{th} parameter of the parameter vector
- $J(\theta)$ – Cost function
- $\hat{z}$ – The estimated value of z
- K- number of states
- $\kappa(x, y)$ – Kernel function
- T – Length of a sequence

Linear Algebra Notation

- tr(A) – Trace of matrix
- det(A) = |A| – Determinant of matrix
- diag(A) = diagonal of matrix
- A^{-1} – Inverse of matrix
- A^{T} – Transpose of matrix
- $\mathbf{I}$ – Identity matrix
- ||x||=$\left||x|\right|_2$ - Euclidean norm, $\sqrt{\sum_{j=1}^{d} x_j^2}$
- $\left||x|\right|_1$ – L1 norm, $\sum_{j=1}^{d}\left|x_j\right|$
- $A_{:,j}$ – j^{th} column of matrix
- $A_{i,:}$ - i^{th} row of matrix
- $A_{i,j}$ – Element (i,j) of matrix

Statistics Notation

- $P(x)$ – Probability of x
- $cov[x]$ – Covariance of x
- $H(X)$ – Entropy of distribution $P(X)$
- $I(X;Y)$ – Mutual information between X and Y
- $\ell(\theta)$ – Log-likelihood function
- μ – Mean
- $p(x)$- Probability density function (pdf)
- ϕ – pdf of standard normal
- σ^2 – Variance
- Σ – Covariance matrix
- μ_{ij} – The mean of the j^{th} feature from class i
- σ_{ij} – The standard deviation of the j^{th} feature from class i