Exam 1 Notes

Inner Product: Calculates the overlap between two vectors

$$a^T b = \|a\|\|b\|\cos(\theta).$$

When $\theta = \pi/2$ the vectors are perpendicular and when $\theta = 0$ the vectors are parallel.

Taylor Series: Approximates a differentiable function to a polynomial at a given point of the original function

$$\mathbb{T}(x, a) = \sum_{n=0}^{\infty} \frac{f^n(x-a)^n}{n!}$$

$$\hat{f}(x) = f(x, z) + \nabla f(z)^T (x - z)$$

$$\nabla f(x, z) = \begin{bmatrix} \frac{\partial f}{\partial x_1}(z) \\ \vdots \\ \frac{\partial f}{\partial x_n}(z) \end{bmatrix}.$$

Affine Function: Inner product with an additional scalar value

$$f(x) = a^T x + b.$$

Linear Functions: Linear functions are of the form

$$\text{Superposition: } f(\alpha x + \beta y) = \alpha f(x) + \beta f(y)$$
$$\text{Homogeneity: } f(\alpha x) = \alpha f(x)$$
$$\text{Additivity: } f(x + y) = f(x) + f(y).$$

Superposition is a combination of homogeneity and additivity.

Linear Regression: Models the relationship between a dependent variable and one or more independent variables, aiming to estimate coefficients that minimize the error between predicted and observed values

$$\hat{y} = x^T \beta + v$$

where x are the regressors, $\hat{y}$ is the prediction, β is an n-vector and v is a scalar.

Norm Of Vector: Measurement of how large a vector or a collection or vectors are

$$\|x\| = x^T x = \sqrt{x_1^2 + \cdots + x_n^2}$$

$$\|x + y\| = \sqrt{\|x\|^2 + 2x^T y + \|y\|^2}$$

$$\|(a, b, c)\| = \sqrt{\|a\|^2 + \|b\|^2 + \|c\|^2}.$$

Triangle Inequality: The norm of a sum of two vectors is no more than the sum of their norms

$$\|x + y\| \leq \|x\| + \|y\|.$$

Root Mean Square (RMS): Typical value of $\|x_i\|$

$$\mathbf{rms}(x) = \frac{\|x\|}{\sqrt{n}}.$$

Chebyshev Inequality: Provides an upper bound on the probability that a random variable deviates from its mean by a certain multiple of its standard deviation

$$\frac{k}{n} \leq \left(\frac{\mathbf{rms}(x)}{a}\right)^2$$

where k is the number of entries in x with absolute value of at least a.

Distance: Calculates the Euclidean distance between two vectors

$$\mathbf{dist}(a, b) = \|a - b\|.$$

Average: Calculates the average value of a vector

$$\mathbf{avg}(x) = \frac{x_1 + \cdots + x_n}{n}.$$

Nearest Neighbor: The vector that is the closest vector to another given vector

$$||x - z_j|| \leq ||x - z_i||, \;\; i = 1, \ldots, n.$$

Standard Deviation: A measure of the dispersion of the values within the vector

$$\mathbf{std}(x) = \frac{\|x - (\mathbf{1}^T x/n)\|}{\sqrt{n}}.$$

J-Clust: J_C is a method of quantifying how good a clustering of data points is for a given cluster

$$J_C = \sum_{i=1}^{N} \sum_{j=1}^{k} \frac{1}{N} \mathbf{min}(\|x_i - z_j\|^2)$$

where z_j is calculated by

$$z_j = \frac{1}{|G_j|} \sum_{i \in G_j}^{n} x_i.$$

K-Means Algorithm: The k-means algorithm clusters data for a given data set into k clusters.

1. *Partition the vectors into k groups:* For each vector $i = 1, \ldots, N$, assign x_i to the group associated with the nearest representative
2. *Update representatives:* For each group $j = 1, \ldots, k$, set z_j to be the mean of the vectors in group j.

Linear Independence: Linear independence and dependence can be summarized with

$$\beta_1 a_1 + \cdots + \beta_k a_k = 0$$

where for a collection of vectors to be linearly independent or dependent one of the following statements must be true.

1. *Linearly Independent:* All values of β must be zero.
2. *Linearly Dependent:* There exists at least one β that is non zero.