02-KNN

Ethan Graham

20th June 2023

Covers Theorem (*Reminder*)

A complicated classification problem cast into high-dimensional space non-linearly is more likely to be linearly separable than in low-dimensional space.

Most things that we want to classify exist on a lower-dimensional manifold of their feature space. For example faces occupying only a small fraction of all possible pictures.

Dimensionality Reduction

The goal is to find a mapping $\mathbf{y}_i = f(\mathbf{x}_i)$ such that $\mathbf{x}_i \in \mathbb{R}^D$ is a high-dimensional data sample and $y_i \in \mathbb{R}^d$ is a low-dimensional representation. We will start by describing a linear approach to this, *i.e.* we want to find $\mathbf{W} \in \mathbb{R}^{D \times d}$ such that $\mathbf{y}_i = \mathbf{W}^T \mathbf{x}_i$

Principle Component Analysis (*PCA*)

The objective is to keep the *important* part of the signal while removing the noise. This can be achieved by finding directions that have large variance. That is, for the j^{th} dimension, we want to maximize

$$var(\{y_i^{(i)}\}) = \frac{1}{N} \sum_{i=1}^{N} (y_i^{(j)} - \overline{y}^{(j)})^2$$

Where $\overline{y}^{(j)}$ is the mean of the dimension of the j^{th} data point after projection.

Variance Maximization in 1 Dimensional space

The variance of the data after projection is

$$var(\{\mathbf{y}_i\}) = \mathbf{w_1^T C w_1}$$

Where $\mathbf{w_1^T w_1} = 1$ and $\mathbf{C}$ is the covariance matrix.

$$\mathbf{C} = \frac{1}{N}\sum_{i=1}^{N}(\mathbf{x}_i - \overline{\mathbf{x}})(\mathbf{x}_i - \overline{\mathbf{x}})^T$$

Ultimately, we are trying to solve

$$\underset{w_1}{max}\, \mathbf{w_1^T C w_1} \text{ subject to } \mathbf{w_1^T w_1} = 1$$

We write the *Lagrangian* for this problem (since it is a constrained optimization problem)

$$L(\mathbf{w}_1, \lambda_1) = \mathbf{w_1^T C w_1} + \lambda_1(1 - \mathbf{w_1^T w_1})$$

$$\frac{\partial L}{\partial \mathbf{w}_1} = 2(\mathbf{C w_1} - \lambda_1 \mathbf{w}_1)$$

The second of which is $= 0$ at the minimum.