Symbols, Patterns and, Signals Notes

paraphrased by Tyler Wright

An important note, these notes are absolutely ***NOT*** *guaranteed to be correct, representative of the course, or rigorous. Any result of this is not the author's fault.*

Contents

1 Data Acquisition

1.1 Analogue to Digital Conversion

There are two steps to this conversion, sampling and quantisation. They can be done in any order.

1.1.1 Nyquist-Shannon Sampling Theorem

If a function f contains no frequencies higher than some $h_{\max}$ hertz, it is completely determined by sampling at points spaced $\frac{1}{2 \cdot h_{\max}}$ apart.

2 Data Characteristics

2.1 Measures of Distance

A valid distance measure $D : A \times A \to \mathbb{R}$ for some data set A has the following properties, it is:

- non-negative,
- reflexive $(D(a, b) = 0 \iff a = b)$,
- symmetric,
- satisfies the triangle inequality $(D(a, b) + D(b, c) \geq D(a, c))$.

2.1.1 Euclidean Distance in $\mathbb{R}^n$ (p-norm distance)

For two vectors x and y in $\mathbb{R}^n$, we have the Euclidean distance D is:

$$D(x, y) := \sqrt[p]{\sum_{i=1}^{n} |x_i - y_i|^p}.$$

2.1.2 Chebyshev Distance in $\mathbb{R}^n$ (∞-norm distance)

For two vectors x and y in $\mathbb{R}^n$, we have the Chebyshev distance D is:

$$D(x, y) := \lim_{n \to \infty} \left(\sqrt[p]{\sum_{i=1}^{n} |x_i - y_i|^p} \right) = \max_{i \in [n]}(|x_i - y_i|).$$

2.1.3 Time Series Distance

Finding the distance between two time series x and y of length n and m (resp.) can be found using Dynamic Time Warping:

$$D_{tw}(x, y) := D(x_1, y_1) + \min\{D_{tw}(x, y'), D_{tw}(x', y), D_{tw}(x', y')\},$$

where D is some numerical distance measure and x' and y' are the time series length $n-1$ and $m-1$ (resp.) corresponding to x and y with the first element removed.

2.1.4 Hamming Distance

When given two strings of the same length, the Hamming distance between them is how many characters differ in the strings at each index.

2.1.5 Edit Distance

When given two strings of any length, the edit distance between them is the smallest number of insertions, substitutions and, deletions that can transform one string into the other (or vice versa).

2.1.6 Wu and Palmer Distance

This measure is based on a hierarchy of word semantics, a graph of relationships between words based on meaning. Using the shortest distance between the words d_1 and the shortest distance from a most specific ancestor to the path d_2 we have the distance measure:

$$D(w_1, w_2) := \frac{2 \cdot d_2}{d_1 + 2 \cdot d_2} - 1.$$

2.2 Summary Statistics

2.2.1 Mean

We take $X = \{x_1, \ldots, x_n\}$ to be a data set. The mean $\bar{X}$ is defined as follows:

$$\bar{X} := \frac{1}{n} \sum_{i=1}^{n} x_i.$$

2.2.2 Standard Deviation and Variance

We take $X = \{x_1, \ldots, x_n\}$ to be a data set. We have that for σ_X, the standard deviation of X, the variance of X is σ_X^2. We define the variance (and thus the standard deviation) as follows:

$$\sigma_X^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{X})^2.$$

2.2.3 Covariance

We take $X = \{x_1, \ldots, x_n\}$ to be a data set consisting of m-dimensional row vectors. We define the covariance matrix:

$$\Sigma = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \mu)^T(x_i - \mu),$$

where μ is the m-dimensional row vector where the j^{th} entry corresponds to the mean of the j^{th} entry of each x_i. This yields a $m \times m$ matrix that is square and symmetric with the variance of the j^{th} entry of each x_i on the j^{th} value on the diagonal.

Eigenvalues and Eigenvectors As the matrix is symmetric we can diagonalise it and find the eigenvalues and eigenvectors. The major axis is the eigenvector corresponding to the largest eigenvalue and the minor axis is the eigenvector corresponding to the smallest value.

2.2.4 Data Normalisation

Data may need to be normalised before we use our distance measures on it. We consider a data set $X = \{x_1, \ldots, x_n\}$ with mean μ and standard deviation σ.

Scaling We map each x_i for $i \in [n]$ as follows:

$$x_i \mapsto \frac{x_i - \min(X)}{\max(X) - \min(X)}.$$

Standardisation We map each x_i for $i \in [n]$ as follows:

$$x_i \mapsto \frac{x_i - \mu}{\sigma}.$$

Scaling to Unit Length We map each x_i for $i \in [n]$ as follows:

$$x_i \mapsto \frac{x_i}{|x_i|}.$$

where $|x_i|$ denotes the magnitude of x_i.

2.2.5 Outliers

A small amount of values significantly different to the remainder of the data set.