Chapter 1

1.1 silhouette coefficient

To calculate the silhouette coefficient of a single example in our dataset, we can apply the following three steps:

1. Calculate the cluster cohesion, $a^{(i)}$, as the average distance between an example, $\boldsymbol{x}^{(i)}$, and all other points in the same cluster.

2. Calculate the cluster separation, $b^(i)$, from the next closest cluster as the average distance between the example, $\boldsymbol{x}^{(i)}$, and all examples in the nearest cluster.

3. Calculate the silhouette, $s^{(i)}$, as the difference between cluster cohesion and separation divided by the greater of the two, as shown here:

$$s^{(i)} = \frac{b^{(i)} - a^{(i)}}{\max\{b^{(i)}, a^{(i)}\}} \tag{1.1}$$

Chapter 2

Activation Functions

2.1 Estimating class probabilities in multiclass classification via the softmax function

The softmax function is a soft form of the argmax function; instead of giving a single class index, it provides the probability of each class. Therefore, it allows us to compute meaningful class probabilities in multiclass settings (multinomial logistic regression).

In softmax, the probability of a particular sample with net input z belonging to the ith class can be computed with a normalization term in the denominator, that is, the sum of the exponentially weighted linear functions:

$$p(z) = \sigma(z) = \frac{e^{z_i}}{\sum_{j=1}^{M} e^{z_j}} \tag{2.1}$$

2.2 Broadening the output spectrum using a hyperbolic tangent

Another sigmoidal function that is often used in the hidden layers of artificial NNs is the hyperbolic tangent (commonly known as tanh), which can be interpreted as a rescaled version of the logistic function:

Chapter 3

Information Theory

3.1

3.1.1 互信息, Mutual Information

离散变量的互信息

在概率论和信息论中，两个随机变量的互信息（mutual Information，MI）度量了两个变量之间相互依赖的程度。具体来说，对于两个随机变量，MI是一个随机变量由于已知另一个随机变量而减少的“信息量”（单位通常为比特）。互信息的概念与随机变量的熵紧密相关，熵是信息论中的基本概念，它量化的是随机变量中所包含的“信息量”。

离散随机变量X和Y的互信息可以计算为：

$$I(X;Y) = \sum_{y \in Y} \sum_{x \in X} p(x,y) \log\left(\frac{p(x,y)}{p(x)\,p(y)}\right), \tag{3.1}$$

其中 $p(x,y)$ 是X和Y的联合概率质量函数，而$p(x)p(y)$ 分别是X和Y的边缘概率质量函数。Mutual Information