KNN

! Assume we have a training dataset D made up of $(x_i)_{i \in [1, n]}$ Samples where $(n = 101)$
The examples are described by a set of features F and any numeric features have been normalised to the range [0, 1]. Each training example is labelled with a Class Label $y_i \in Y$.

Objective: classify an unknown example q.
(q can be called a query)

For Each $x_i \in D$ we can calculate the distance between q & x_i as follows.

$$d(q, X_i) = \sum_{f \in F} w_f \, \delta(q_f, X_{if})$$

This is a Summation over all the features in F with w_f the weight for each feature

$$\delta(q_f, X_{if}) = \begin{cases} 0, & f \text{ discreet and } q_f = X_{if} \\ 1, & f \text{ discreet and } q_f \neq X_{if} \\ |q_f - X_{if}|, & f \text{ continuous} \end{cases}$$

The KNN are selected based on this distance metric.

Votes (weighted distance)

$$Vote(y_j) = \sum_{c=1}^{K} 1/d(q, x_c)^p \cdot g(y_j, y_c)$$

$$g(a,b) = 1 \text{ if } a_{label} = b_{label} \text{ else } 0$$
