## KNN

KNN can be used both in regression and classification. It is a method of lazy learning. This method is based on the assumption that neighbourhoods have some similarity.

### 1. Form

- Regression

$$\hat{f}(x) = \frac{1}{K}\Sigma_{i \in N_{K(x)}} y_i$$

- Classification

  Asign the label of the majority of K neighbours.

  *Note, it has a non-linear boundary.*

### 2. Additionals

- K is a tuning parameter in KNN

  Higher K, boundary/fit line is smoother, bias is larger, variance is smaller.

- KNN can fail in high dimensions, because it becomes difficult to gather K observations close to target point due to the spartial sparsity.

  *Note: This problem is actually called **curse of dimensionality**. We could solve it with dimension reduction or variable selection.*

- What affects KNN?
  - K
  - How to compute distance
  - Whether the weighted strategy applies