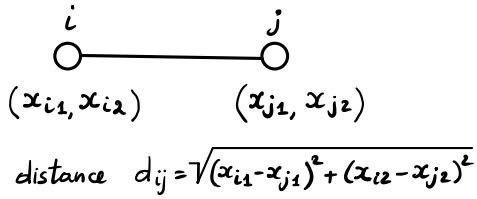
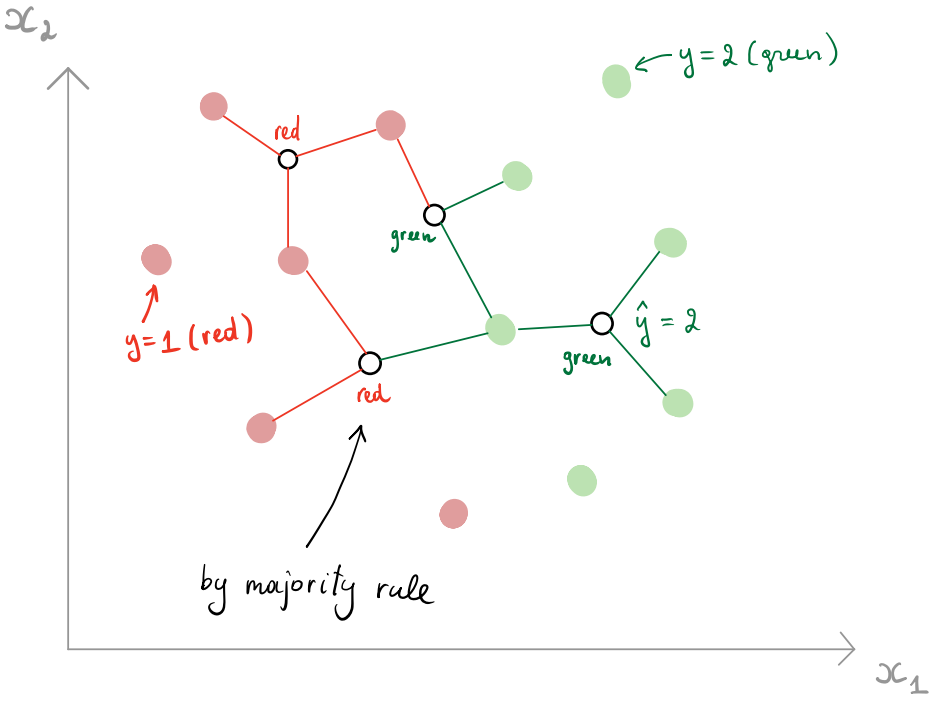
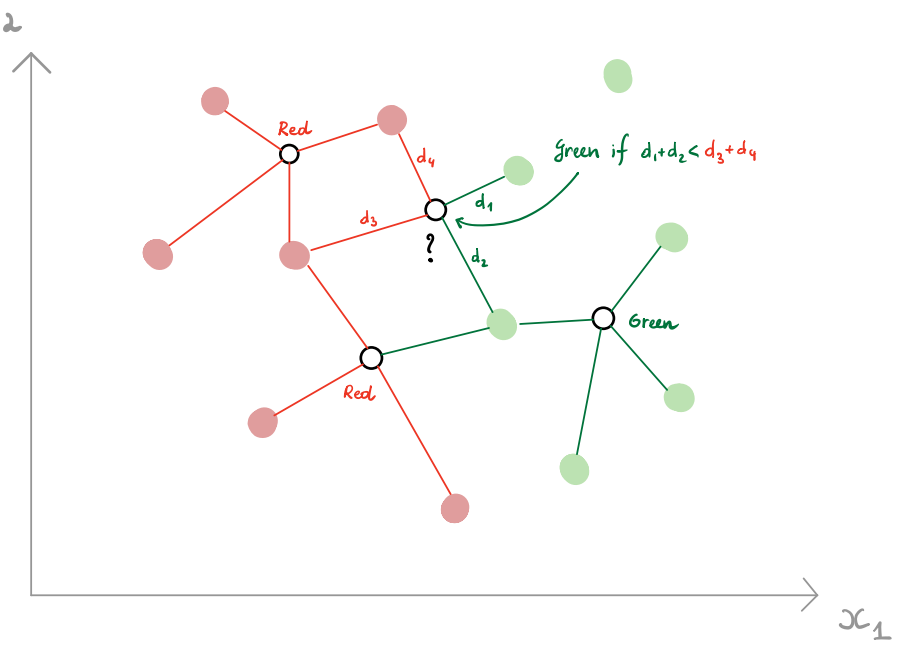
KNN classification



K = 3

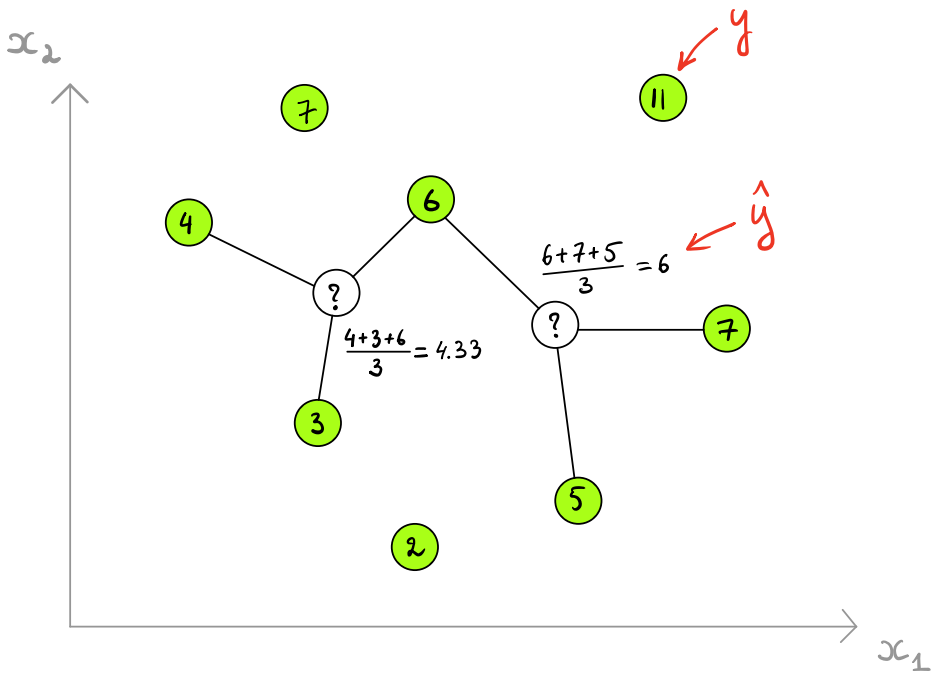


K = 4



KNN Regression

K = 3



Basic k-mean:

$$\hat{y} = \frac{1}{k} \sum_{i=1}^{k} y_i$$

Weighted k-mean:

$$\hat{y} = \frac{1}{k} \sum_{i=1}^{k} w_i y_i \quad \text{where} \quad w_i = \frac{\frac{1}{d_i}}{\sum_{j=1}^{k} \frac{1}{d_j}} \qquad \sum_i w_i = 1$$

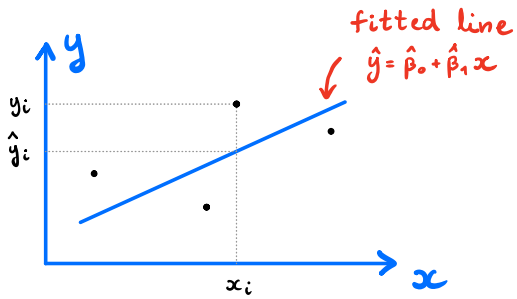
Classification: [1, 2, 2] [0.8, 0.2, 0.2]

use numpy.bincount()

Linear regression:

- Single predictor $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$

regression coefficients; random error; response; predictor



- Multiple predictors $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \varepsilon_i$

Matrix form $\vec{y} = X\vec{\beta} + \vec{\varepsilon}$

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & x_{12} & \cdots \\ 1 & x_{21} & x_{22} & \cdots \\ \vdots & \vdots & \vdots & \\ 1 & x_{n1} & x_{n2} & \cdots \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

responses; design matrix; regression coeficients; errors

Prediction $\hat{\vec{y}} = X^{(test)}\hat{\vec{\beta}}$

$$\hat{\vec{\beta}} = \left(\left(X^{(train)}\right)^T X^{(train)}\right)^{-1} \left(X^{(train)}\right)^T \vec{y}$$

$$\begin{pmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \vdots \\ \hat{y}_m \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & x_{12} \\ 1 & x_{21} & x_{22} \\ \vdots & \vdots & \vdots \\ 1 & x_{m1} & x_{m2} \end{pmatrix}$$