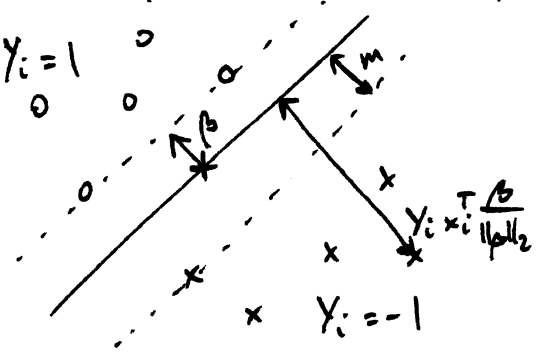
1.6 Other kernel machines

The least squares loss of KRR seems appropriate when the response is cts. We now consider the case where $Y \in \{-1, 1\}^n$, so the responses are class labels.

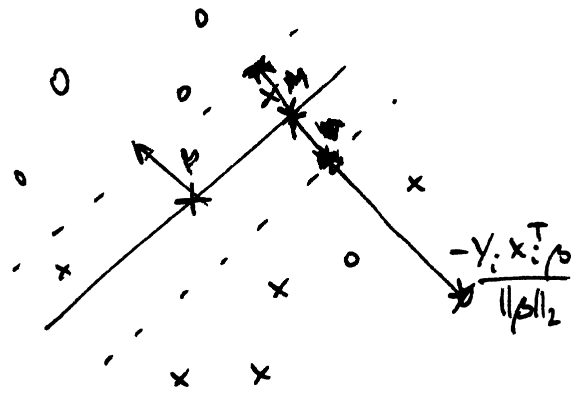
1.6.1 The support vector machine

Suppose $\{x_i\}_{i: Y_i = 1}$ and $\{x_i\}_{i: Y_i = -1}$ are separable by a hyperplane through the origin, i.e. $\exists \beta \in \mathbb{R}^p$ st. $Y_i x_i^T \beta > 0 \ \forall i$.



There is an infinite number of planes that separate the two classes. One approach is to pick the plane such that maximises the margin between the two classes. This is given by the optimisation problem

$$\max_{\beta \in \mathbb{R}^p, M \geq 0} M \quad \text{subj to} \quad \frac{Y_i x_i^T \beta}{\|\beta\|_2} \geq M.$$



We can replace the constraint $\frac{Y_i x_i^T \beta}{\|\beta\|_2} \geq M$ with a penalty for how far over its margin boundary x_i is. The penalty should be zero for those points on the correct side of their margin boundary.

Two natural choices for this penalty are

$$\lambda \sum_{i=1}^n \left(M - \frac{Y_i x_i^T \beta}{\|\beta\|_2}\right)_+$$

$$\lambda \sum_{i=1}^n \left(1 - \frac{Y_i x_i^T \beta}{M \|\beta\|_2}\right)_+$$

The second of these leads to a tractable optimisation problem. Replacing $\max M$ with $\min \frac{1}{M^2}$ and adding the penalty, we get

$$\min_{M \geq 0, \beta \in \mathbb{R}^p} \frac{1}{M^2} + \lambda \sum_{i=1}^n \left(1 - \frac{Y_i x_i^T \beta}{M \|\beta\|_2}\right)_+$$