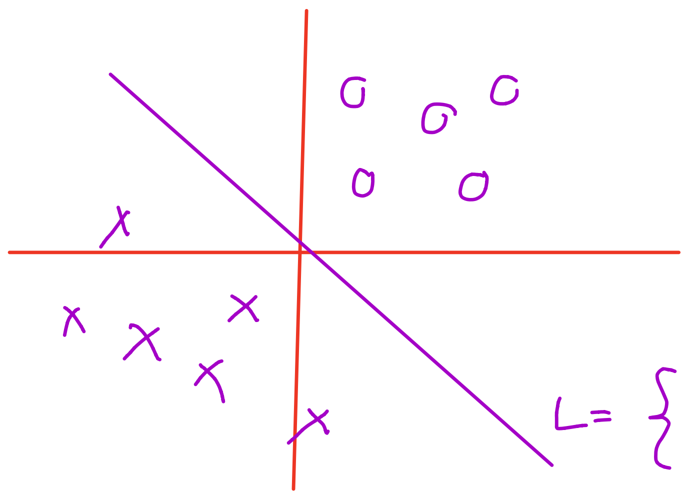
Maximum Margin Hyperplane



$L = \{x : f(x) = 0\}$

$$f(x) = \beta_0 + x^T\beta$$

If the data are perfectly seperable then there exists a unique function that acheives this.

$$(\beta_0, \beta) = \underset{\{\beta_0, \beta \,:\, y_i(x_i^T\beta + \beta_0) \geq 1\}}{\arg\max} \frac{1}{\|\beta\|}$$

only satisfied if perfect classification

$$= \arg\max \frac{1}{\|\beta\|}$$

$\|\beta\|$

$$\left\{\beta_0, \beta : \frac{y_i(x_i^T\beta + \beta_0)}{\|\beta\|} \geq \frac{1}{\|\beta\|}\right\}$$

Goal: find the largest value s.t. the perfectly classified points are furthest from the hyperplane.

Convex Optimization

Looking to solve

$$\underset{\{\beta_0, \beta : y_i(x_i^T\beta + \beta_0) \geq 1\}}{\arg\min} \quad \|\beta\|^2$$

Constrained optimization problem.

Using Lagrange multipliers

$$L_p = \frac{1}{2}\|\beta\|^2 - \alpha_1 (y_1(x_1^T\beta + \beta_0) - 1)$$
$$- \alpha_2 (y_2(x_2^T\beta + \beta_0) - 1)$$
$$\vdots$$
$$- \alpha_N (y_N(x_N^T\beta + \beta_0) - 1)$$

+ constraints to be specified

$$\frac{d L_p}{\partial \beta} = \frac{\partial}{\partial \beta}\left\{ \frac{1}{2}\|\beta\|^2 - \sum \alpha_i \left(y_i (x_i^T \beta - \beta_0) - 1 \right) \right\}$$

$$\Rightarrow \quad \beta = \sum_{i=1}^{N} \alpha_i y_i x_i \qquad\qquad 0 = \sum_{i=1}^{N} \alpha_i y_i$$

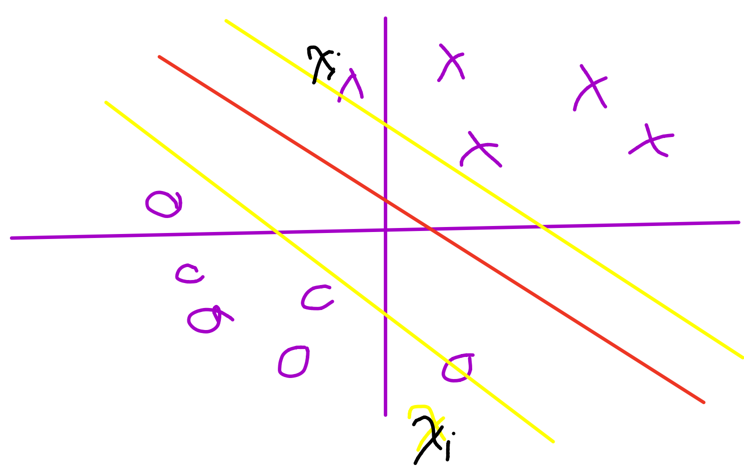
The dual Lagrangian can be written

$$L_p = \sum \alpha_i - \frac{1}{2} \sum^{N} \sum^{N} \alpha_i \alpha_k \, y_i y_k \, x_i^T x_k$$

these constraints on the KKT conditions

$$\alpha_i \left[(y_i (x_i^T \beta + \beta_0 - 1) \right] = 0$$

if $\alpha_i > 0$ then $y_i (x_i^T \beta + \beta_0) = 1.$



these touching points are going to be the only contributors to $\sum \alpha_i y_i x_i$

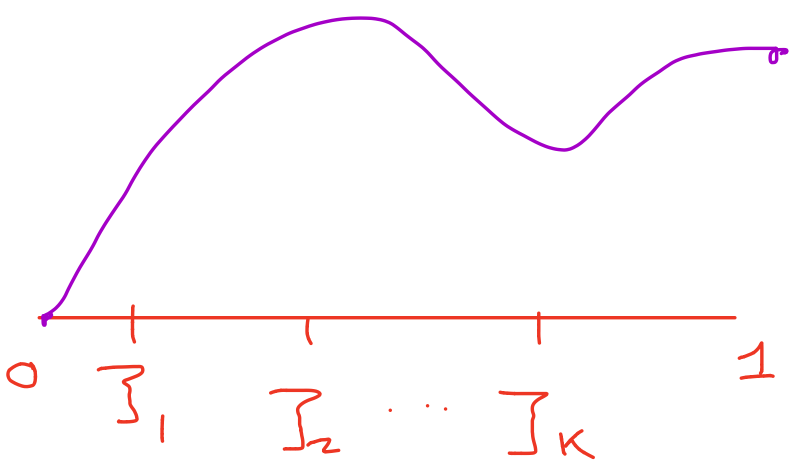
Moreover if $y_i(x_i^T\beta+\beta_0)$ is off the margin and do not contribute to the α_i in the sum $\sum \alpha_i y_i x_i$.

Basis Expansions

$Y=\sum_{i=1}^{N} \beta_i x_i + \beta_0$ replaced with $Y=\sum_{i=1}^{N} \beta_i h(x_i)+\beta_0$

One good choice of functions $h(x)$ are called

splines.



Piecewise cubic polynomials on the intervals (ξ_i, ξ_{i+1})

such that $\hat{f} = \sum_{i=1}^{m} h_i(x_i) \in C^2$