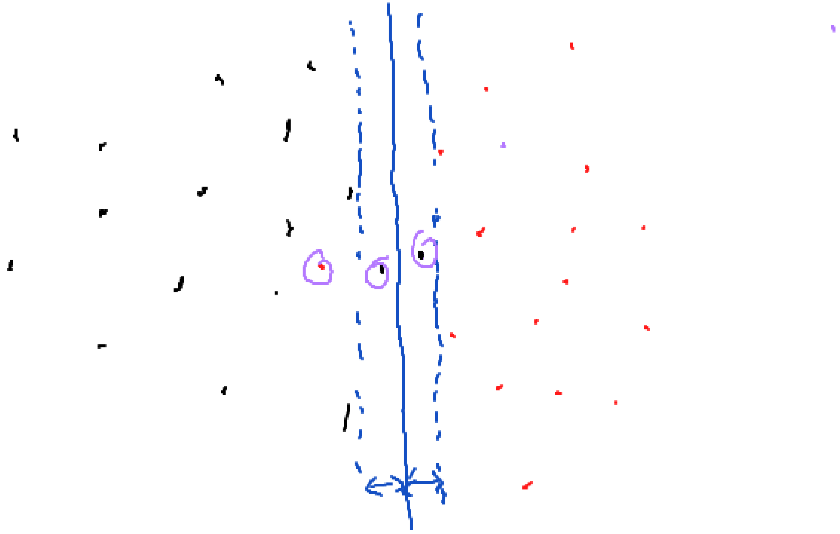


$$f(x) = \overset{\in\mathbb{R}}{\beta_0} + \overset{\in\mathbb{R}^p}{\beta_1} x \qquad \text{classification : } \operatorname{sgn} f(x)$$

it turns out that f is found by fitting parameters α_i, β_0, with

SVM (support vector machine)

$$f(x) = \sum_{i=1}^{n} \alpha_i \underset{\in\{-1,1\}}{y_i} \langle x, x_i \rangle + \beta_0$$

~~A~~ the algorithm (not discussed here) may be generalised to allow functions of the form

$$f(x) = \sum_{i=1}^{n} \alpha_i y_i \underbrace{K(x,x_i)}_{\text{symmetric semi positive definite}} + \beta_0$$

i.e. $\sum_i \sum_j K(z_i, z_j)\, c_i c_j \geq 0$ for all c_i, c_j, z_i, z_j

$$K(x,y) = K(y,x)$$

if $K(z_i, z_j) = \langle z_i, z_j \rangle$, then $\sum_i \sum_j \langle z_i, z_j \rangle c_i c_j = \sum_i \sum_j \langle z_i c_i, z_j c_j \rangle =$

$$= \left\langle \sum_i c_i z_i, \sum_j c_j z_j \right\rangle = \left\| \sum_i c_i z_i \right\|^2 \geq 0$$

Examples

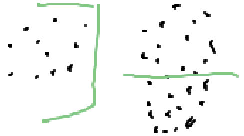
$$K(x,x') = (1 + \langle x, x' \rangle)^d$$

$$K(x,x') = \exp(-\gamma \|x - x'\|^2)$$

Remarks: SVMs ~~work~~ well for large p, even if n is not too large

p = number of predictors

n = number of samples

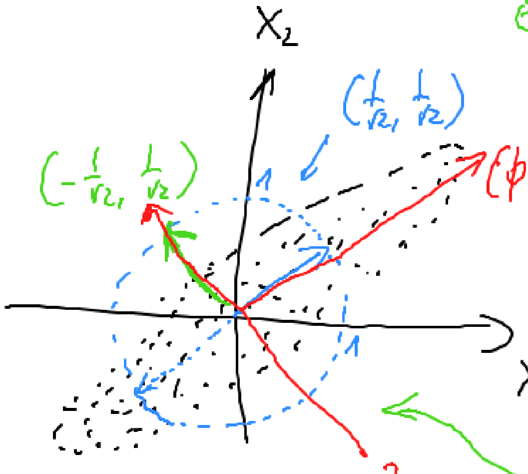


Orthonormal collection of vectors

$$\{(\phi_{11}, \phi_{21}, \dots, \phi_{p1}), \dots, (\phi_{1p}, \phi_{2p}, \dots, \phi_{pp})\}$$

this defines a new coordinate system in which 'the most important' predictor is the first coordinate, and so on.

Example



$$(X_1, X_2) \circ \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right) = \frac{1}{\sqrt{2}}(X_1 + X_2)$$

$$(X_1, X_2) \circ \left(-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right) = \frac{1}{\sqrt{2}}(-X_1 + X_2)$$

<u>Application</u>: reducing the dimension –

– instead of having p predictors, we can have $m < p$ predictors.

The first assignment should be finished until 17^{th} December (i.e., you should send the report and source code)

3rd assignment: find some data that needs some processing (i.e., not from the library) and do either SVM or k-means clustering