Support Vector Classifier

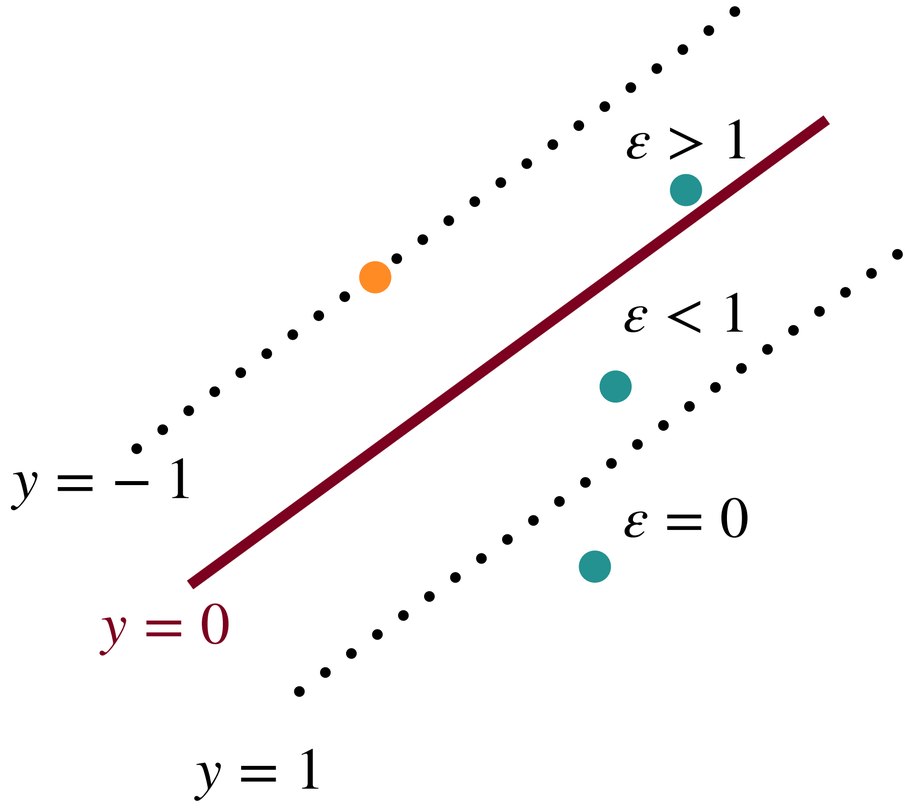
$$\max_{\beta_0,\beta_1,\ldots,\beta_p,\varepsilon_1,\varepsilon_2,\ldots,\varepsilon_n} M$$

subject to:

$$\|\beta\| = 1$$

$$y_i(\beta_0 + \beta^T x_i) \geq M(1 - \varepsilon_i)$$

$$\varepsilon_i \geq 0, \sum_{i=1}^{n} \varepsilon_i \leq C$$



C is the tuning parameter/penalty on error:

$C = 0$ implies maximal margin hyperplane (superposed it exists)

$C > 0$ is the total violations to the margin that we can tolerate

$\Longrightarrow$ max C observations can be on the wrong side of hyperplane

Support Vector Classifier

C: penalty on error

	Regularization	Margins	Bias/Variance
Small C	more	wider	prone to underfitting
Large C	less	narrower	prone to overfitting