c defines the violation allowed. But in practice c is a variable choosen through $\boxed{\text{cross-validation}}$*.

c controls the bias-variance tradeoff..

c ↓ low bias, high variance
C ↑ high bias, low variance.

Support vector machines

The examples that we took were for 2-class model and for linear factors. what if we had non-linear factors.

$$X_1, X_2, \ldots \ldots X_p$$

and

$$X_1^2, X_2^2 \ldots \ldots X_p^2.$$

we would then maximize

$$\max_{\beta_1, \beta_2, \ldots \beta_p - \beta_1^2 \ldots \beta_p^2} M.$$

Subject to

$$y_i \left(\beta_0 + \sum_{j=1}^{p} \beta_{j1} X_{ij} + \sum_{j=1}^{p} \beta_{j2} X_{ij}^2 \right) \geq M(1-\epsilon_i)$$

$$\sum_{i=1}^{n} \epsilon_i \leq c, \quad \epsilon_i \geq 0, \quad \sum_{i=1}^{p} \sum_{k=1}^{2} \beta_{jk}^2 = 1.$$

This would yeild a non-linear decision boundry.

so support vector machine is an enlargement of support vector classifier.