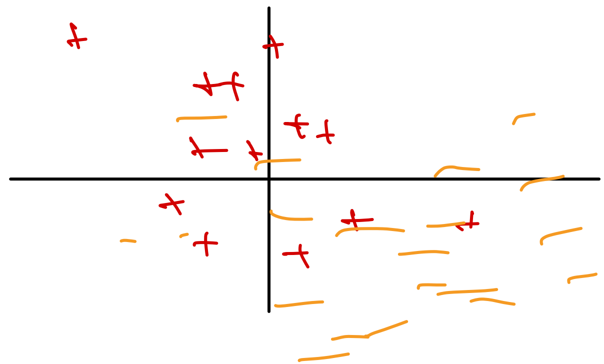
Nonseparable case



Max margin classifier can be sensitive to the addition of new data, and for nonseparable data, it doesn't exist.

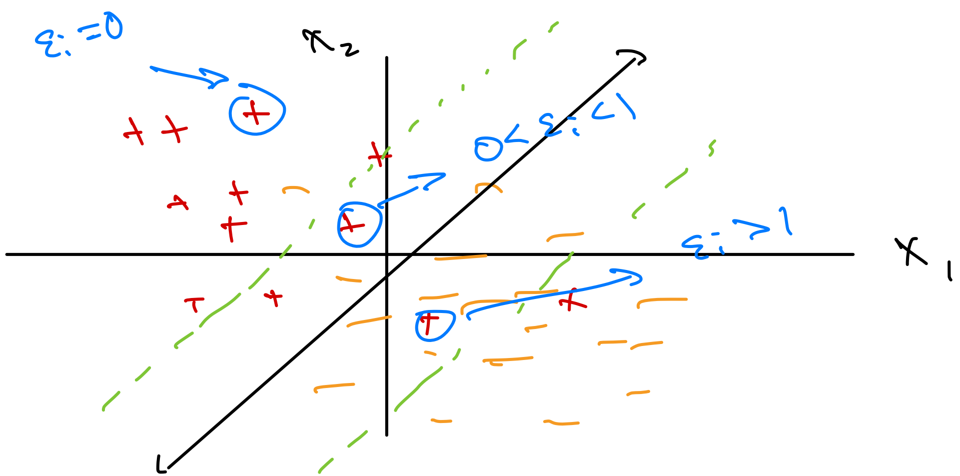
A soft margin or support vector classifier (soft = some data will be misclassified) is calculated via:

- $\underset{\beta_0, \underline{\beta}, \varepsilon_1, \ldots, \varepsilon_n}{\text{Maximize}} \quad M$
- subject to $\|\underline{\beta}\|_2 = 1$ and $y_i(\beta_0 + \underline{\beta}^T \underline{x}_i) \geq M(1 - \varepsilon_i)$

  and $\sum_{i=1}^{n} \varepsilon_i \leq \underline{C}$ with $\varepsilon_i \geq 0 \ \forall i$

$\boxed{\text{Note}}$ Separable case handled with $\varepsilon_1 = \cdots = \varepsilon_n = 0$

- ε_i are slack variables where
  - $\varepsilon_i = 0$ then $(\underline{x}_i, y_i)$ is correctly classified + outside of margin
  - $0 < \varepsilon_i \leq 1$ " " " " " but inside margin
  - $\varepsilon_i > 1$ " " is misclassified



The param C chosen
↳ by cross-validation

Defining the support vectors

Can still write program in Lagrangian coords

$$\frac{1}{2}\|\underline{\beta}\|_2^2 + C\sum_{i=1}^n \varepsilon_i - \sum_{i=1}^n \alpha_i \left(y_i(\beta_0 + \underline{\beta}^T \underline{x}_i) - (1-\varepsilon_i)\right) - \sum_{i=1}^n \mu_i \varepsilon_i$$

set derivs $= 0$ yields:

$$\underline{\beta} = \sum_{i=1}^n \alpha_i y_i \underline{x}_i \qquad \sum_{i=1}^n \alpha_i y_i = 0$$

$\Rightarrow$ solution has form

$$f(\underline{x}) = \beta_0 + \underline{\beta}^T \underline{x} = \underline{\beta_0 + \sum_{i=1}^n \alpha_i y_i \underline{x}_i^T \underline{x}}$$

f is still a linear classifier + usually α_i are 0,

when $\alpha_i \neq 0$ then $\underline{x}_i$ is a support vector.