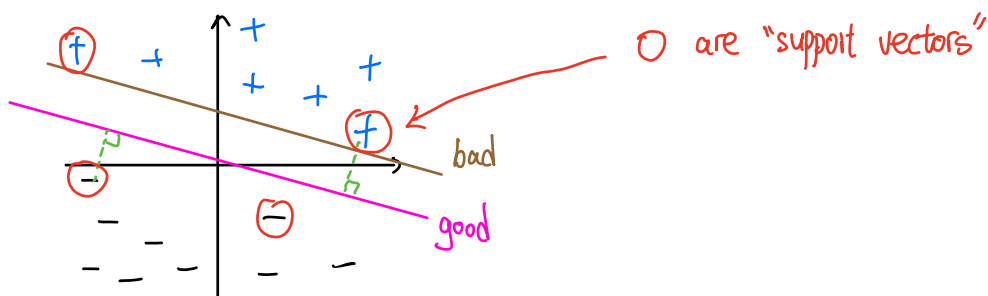


Support Vector Machine

- competitor to logistic regression
 - do binary classification
 - learn linear decision boundaries
 - + often paired with kernels for efficiency reasons

intuition: Consider linearly separable data



SVM chooses decision boundary where even closest point is far from boundary.

Far away points don't affect the decision boundary, which is influenced by closest points.

Kernel SVM

we also define score = $\sum_{i=1}^n a_i k(x^{(i)}, x)$ for input x

a_i for non-support vectors will be 0

therefore, evaluating kernel SVM takes time proportional to

support vectors < number of training examples

Fact: with no kernel, SVM is solving the minimization problem

$$[z]_+ = \begin{cases} z & \text{if } z > 0 \\ 0 & \text{else} \end{cases}$$

$$L(w) = \left(\sum_{i=1}^n \underbrace{[1 - y^{(i)} w^T x^{(i)}]_+}_{\text{margin}} \right) + \underbrace{\lambda \|w\|^2}_{L2 \text{ regularization}}$$

$[1 - \text{margin}]_+$ is called "hinge loss"

compare with logistic regression

