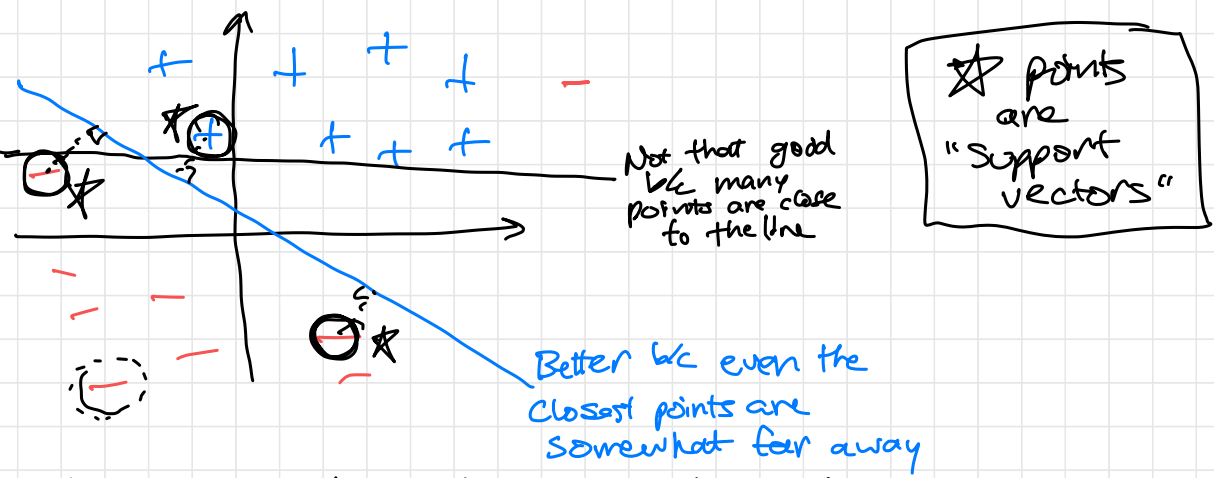
2/14/2023 Support vector machines (SVM)

- Competitor to logistic Regression
  - Also doing binary classification
  - Also learn linear decision boundary
  - * Often paired with kernels b/c efficiency reasons

Intuition Consider linearly separable data



☆ points are "support vectors"

Idea of SVM: Choose decision boundary where even closest point is far from boundary

Faraway points don't affect decision boundary
(If you move one, doesn't change whether close points are close to boundary)
→ Only support vectors (close points to boundary) influence the boundary

Kernel SVM

We also define score = $\sum_{i=1}^{n} \alpha_i K(x^{(i)}, x)$ for input x

α_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:

$$[z]_+ = \text{"positive part" of } z = \begin{cases} z & \text{if } z > 0 \\ 0 & \text{if } z \le 0 \end{cases}$$

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

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

Compare with Logistic Regression