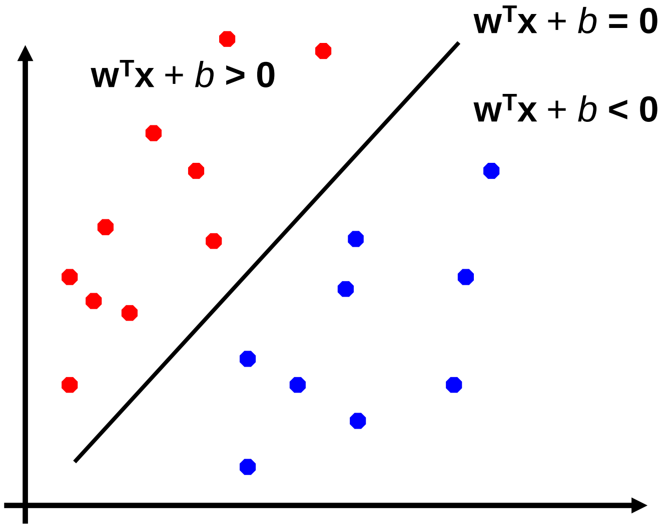
(support vector machine)
(SVM)

Perceptron Revisited: Linear Separators

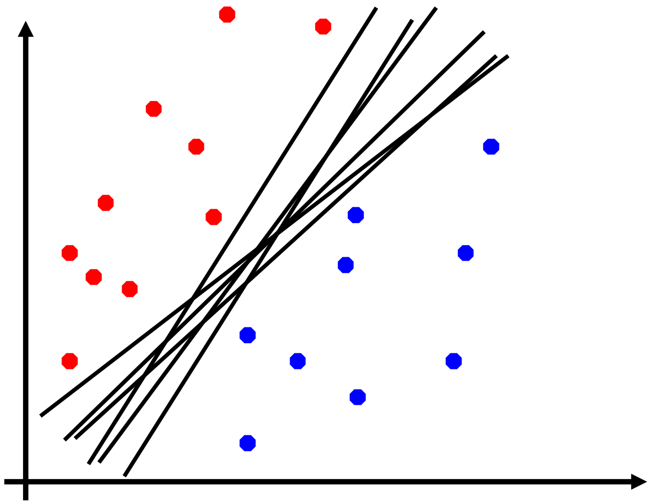
- Binary classification can be viewed as the task of separating classes in feature space:



$f(\mathbf{x}) = \text{sign}(\mathbf{w}^T\mathbf{x} + b)$

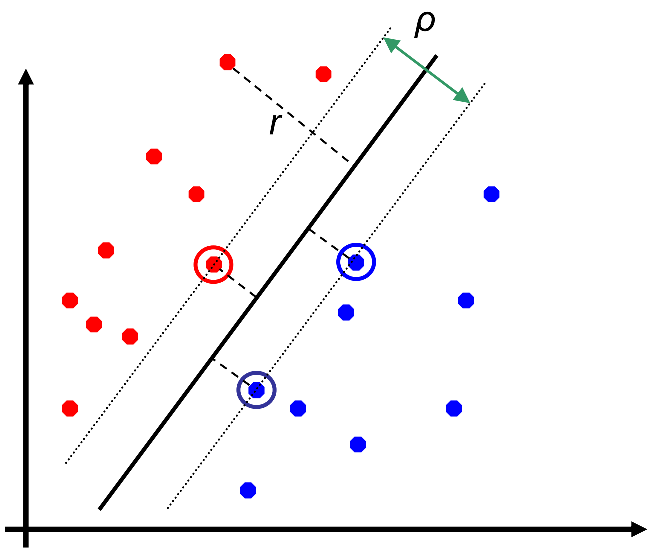
Linear Separators

- Which of the linear separators is optimal?



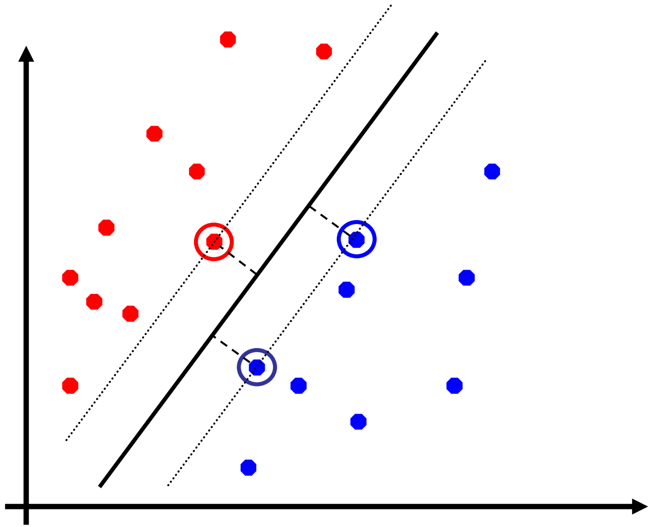
Classification Margin

- Distance from example $\mathbf{x}_i$ to the separator is $r = \frac{\mathbf{w}^T \mathbf{x}_i + b}{\|\mathbf{w}\|}$
- Examples closest to the hyperplane are ***support vectors***.
- ***Margin*** *ρ* of the separator is the distance between support vectors.

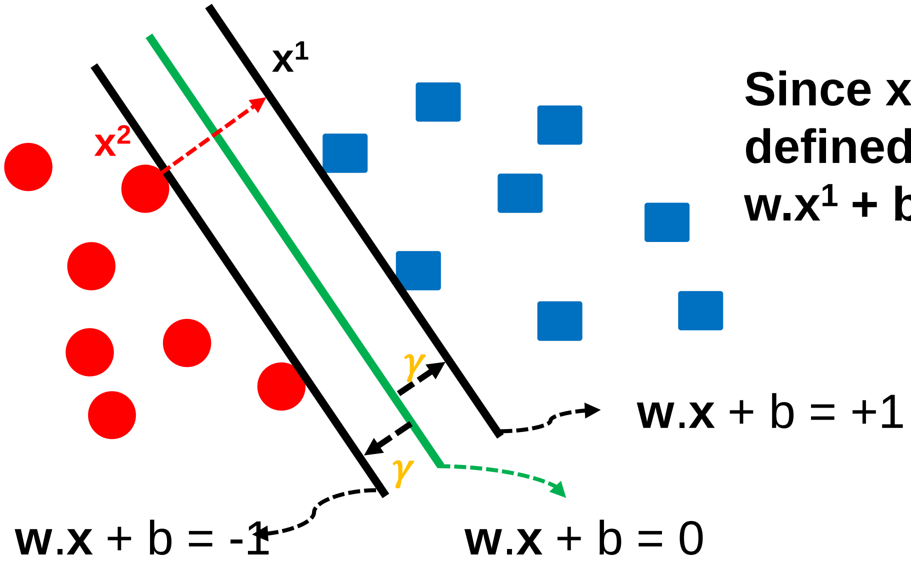


Maximum Margin Classification

- Maximizing the margin is good according to intuition and PAC theory.
- Implies that only support vectors matter; other training examples are ignorable.



The Objective of SVM



Since x^1 is on the hyperplane defined by w.x + b = +1, we know that w.x^1 + b = 1. If we substitute for x^1:

Towards Learning an SVM

- How to learn an SVM $\boldsymbol{h_\theta(x) = \theta^T x}$, where $\boldsymbol{\theta}$ = $[\boldsymbol{\theta_0}, \ldots, \boldsymbol{\theta_m}]$ and $\boldsymbol{x} = [\boldsymbol{x_0}, \ldots, \boldsymbol{xm}]$?
  - Say, by minimizing ***Mean Squared Error (MSE)***. That is:

$$\text{minimize} \frac{1}{2n} \sum_{i=1}^{n} \left(y'^{(i)} - y^{(i)}\right)^2$$

$$\equiv$$

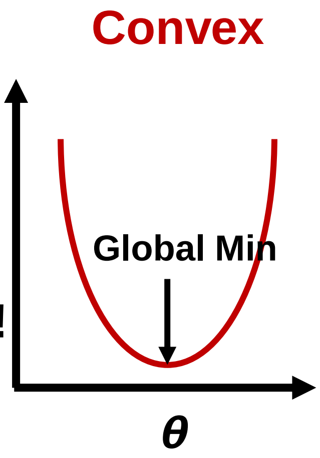
$$\underset{\theta}{\text{minimize}} \frac{1}{2n} \sum_{i=1}^{n} \left((g(\theta^T x))^{(i)} - y^{(i)}\right)^2$$

$$\equiv$$

$$\underset{\theta}{\text{minimize}}\, \boldsymbol{J(\theta)}$$

Unfortunately, if we plot this cost function, it will turn out to be "**non-convex**

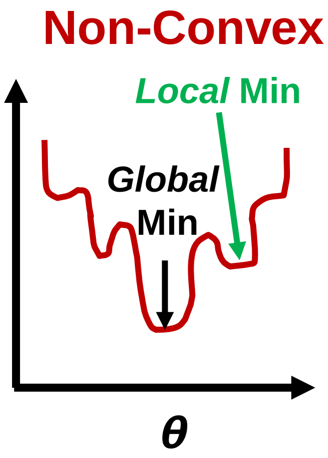
Easy for Gradient Descent to locate!



Towards Learning an SVM

- How to learn an SVM $\boldsymbol{h_\theta(x)} = \boldsymbol{\theta^T x}$, where $\boldsymbol{\theta}$ = $[\boldsymbol{\theta_0}, \ldots, \boldsymbol{\theta_m}]$ and $\boldsymbol{x} = [\boldsymbol{x_0}, \ldots, \boldsymbol{xm}]$?
  - Say, by minimizing ***Mean Squared Error (MSE)***. That is:

Gradient Descent might get stuck at a *local* min and fail to locate the *global* min!



$$\text{minimize} \frac{1}{2n} \sum_{i=1}^{n} \left(y'^{(i)} - y^{(i)}\right)^2$$

$$\equiv$$

$$\underset{\theta}{\text{minimize}} \frac{1}{2n} \sum_{i=1}^{n} \left((g(\theta^T x))^{(i)} - y^{(i)}\right)^2$$

$$\equiv$$

$$\underset{\theta}{\text{minimize}}\, \boldsymbol{J(\theta)}$$

Unfortunately, if we plot this cost function, it will turn out to be “**non-convex**

- SVM Lab time

Question ?