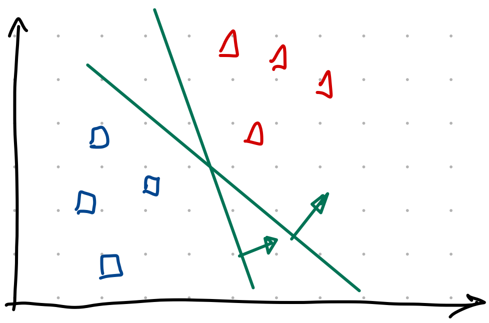
Let's finish the rest of LR

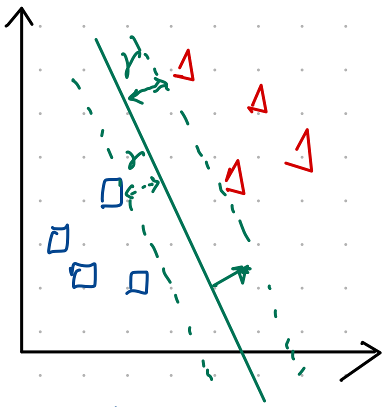
- Programing Assignment 5 (LR in 1D)
- With MLE, we find $\vec{w}$ that minimizes the square loss
- Closed-form solution of LR (normal equation)
- Gradient descent vs. normal equation

Support Vector Machine (SVM)

- Perceptron: find a hyperplane if its exists
- SVM: find the maximum margin separating hyperplane.

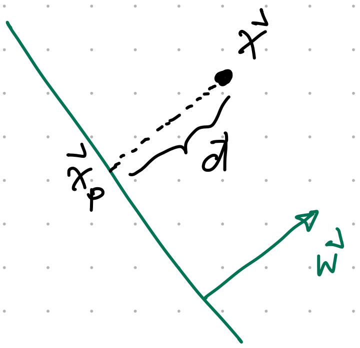
$y \in \{+1, -1\}$





The margin γ is the distance from the hyperplane to the closest points in either class

Margin: $\mathcal{H}_{\vec{w}, b} = \{ \vec{x} \mid \vec{w}^T \vec{x} + b = 0 \}$



$\vec{x}_p = \vec{x} - \vec{d}$

$\vec{w}^T \vec{x}_p + b = 0 \Rightarrow \vec{w}(\vec{x} - \overbrace{\vec{d}}^{= \alpha \vec{w}}) + b = 0$

$\Rightarrow \vec{w}(\vec{x} - \alpha \vec{w}) + b = 0$

$\Rightarrow \alpha = \dfrac{w^T x + b}{w^T w}$

$$\vec{d} = \underbrace{\frac{w^T x + b}{w^T w}}_{\alpha} \vec{w} \Rightarrow \|d\|_2 = \sqrt{d^T d} = \sqrt{\alpha^2 w^T w} = \alpha \sqrt{w^T w}$$

$$= \frac{w^T x + b}{w^T w} \sqrt{w^T w}$$

$$= \frac{w^T x + b}{\sqrt{w^T w}} = \frac{w^T x + b}{\|w\|_2}$$

$$\Rightarrow \boxed{\gamma(\vec{w}, b) = \min_{i \in \{1,2,\ldots,n\}} \frac{\vec{w}^T \vec{x} + b}{\|\vec{w}\|_2}}$$

margin

The goal of SVM is to find

$$\boxed{\vec{w}, b = \max_{\vec{w}, b} \gamma(\vec{w}, b)}$$ objective

subject to

$$\boxed{\forall_i,\ y_i(\vec{w}^T \vec{x}_i + b) \geq 0}$$ constraint

- If we plug in the definition of γ

$$\Rightarrow \vec{w}, b = \max_{\vec{w}, b} \min_{\forall i} \frac{|\vec{w}^T \vec{x}_i + b|}{\|\vec{w}\|_2} \quad \text{subject to} \quad \forall_i,\ y_i(\vec{w}^T \vec{x}_i + b) \geq 0$$

$$\Rightarrow \vec{w}, b = \max_{\vec{w}, b} \frac{1}{\|\vec{w}\|_2} \min_{\forall i} |\vec{w}^T \vec{x}_i + b| \quad \text{subject to} \quad \forall_i,\ y_i(\vec{w}^T \vec{x}_i + b) \geq 0$$

- Because the hyperplane is scale invariant

$$\Rightarrow \underbrace{\min_{\forall i} |\vec{w}^T \vec{x}_i + b| = 1}_{\text{another constraint}}$$

$$\Rightarrow \quad \vec{w}, b = \max_{\vec{w}, b} \frac{1}{\|\vec{w}\|_2} = \min_{\vec{w}, b} \|\vec{w}\|_2 = \min_{\vec{w}, b} \vec{w}^T \vec{w}$$

subject to $\min_{\forall i} |\vec{w}^T \vec{x}_i + b| = 1$ — Constraint A

$\forall i,\ y_i (\vec{w}^T \vec{x}_i + b) \geq 0$ — Constraint B

A and B is true if and only if $\forall i,\ y_i (\vec{w}^T \vec{x}_i + b) \geq 1$.

The Final Formulation:

$$\vec{w}, b = \underbrace{\min_{\vec{w}, b} \vec{w}^T \vec{w}}_{\text{objective}} \quad \text{subject to} \quad \underbrace{\forall i\ y_i (\vec{w}^T \vec{x}_i + b) \geq 1}_{\text{new constraint.}}$$