Week 5

9/23/25

- Made it! Only two problems/week from here on
Lost	Trouble	Difficult	Understood Everything
$\underbrace{1}_{\text{Let's chat!}}$	$\underbrace{6}_{\text{Let's chat}}$	10	3
- OH Changes: Tomorrow → Thursday 12:30-2:00
- Pytorch learning curve
- Student discussions! Snacks!

Review

Classification e.g.,

- spam vs ham
- hand written digits
- breast cancer
- diabetes

Focus on binary (easier visualizations)

① Linear + Sigmoid $f(x) = \sigma(\langle w, x\rangle)$

$$= \frac{1}{1+e^{-\langle w, x\rangle}}$$

② Binary Cross Entropy

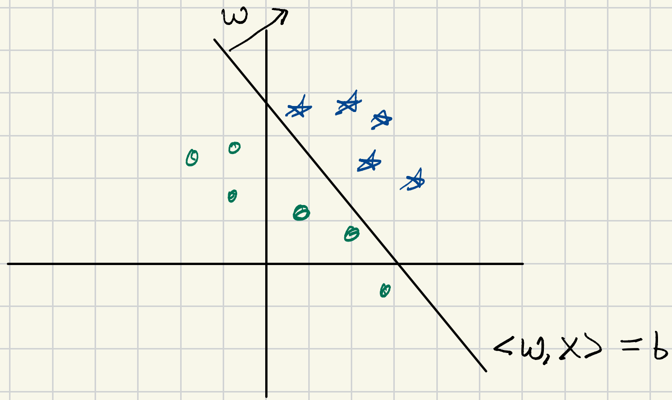
$$y \log(\sigma(\langle w, x\rangle)) + (1-y)\log(1-\sigma(\langle w, x\rangle))$$

③ Gradient Descent

$$\nabla \mathcal{L}(w) = X^T(\sigma(Xw) - y)$$

$$f(x) = \sigma(\langle w, x\rangle) \geq \tau$$

$\iff \langle w, x\rangle \geq b$ where $\tau = \sigma(b)$



In $\mathbb{R}^2$, w, $w^\perp$ form orthonormal basis

$$x = \alpha w + \beta w_\perp$$

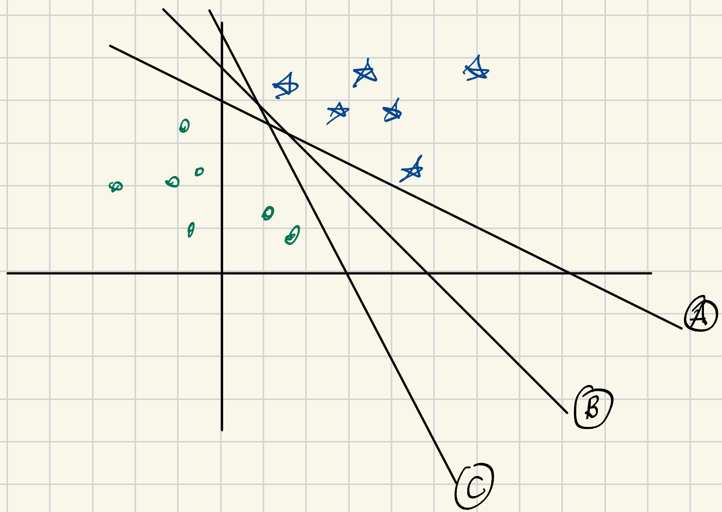
$$\langle x, w\rangle = \alpha\langle w, w\rangle + \beta\langle w_\perp, w\rangle$$
$$= \alpha$$

Support Vector Machines

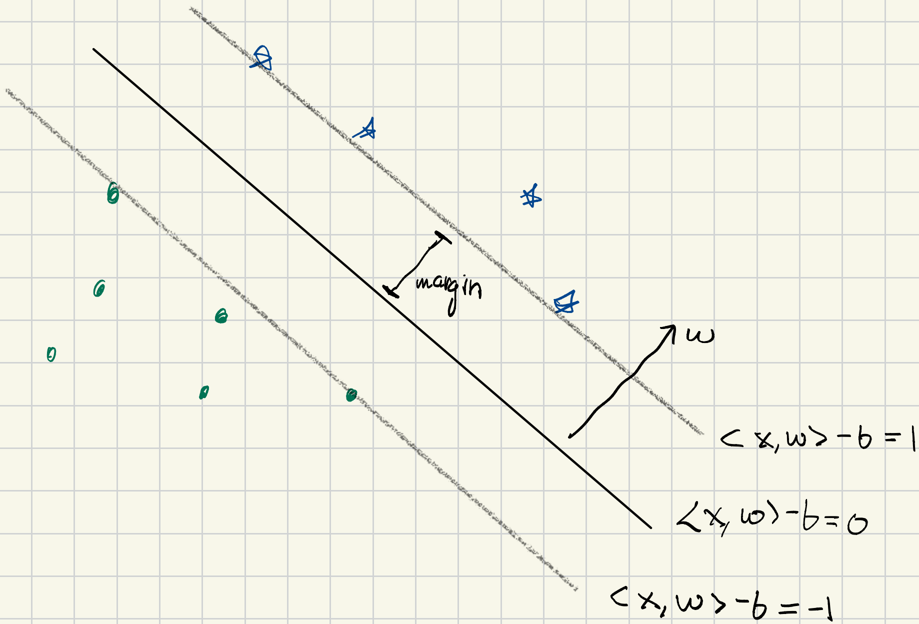
Assumption : linearly separable (extra features help)

$\underline{x}^{(i)} \in \mathbb{R}^d$ $\quad$ $y^{(i)} \in \{-1, 1\}$

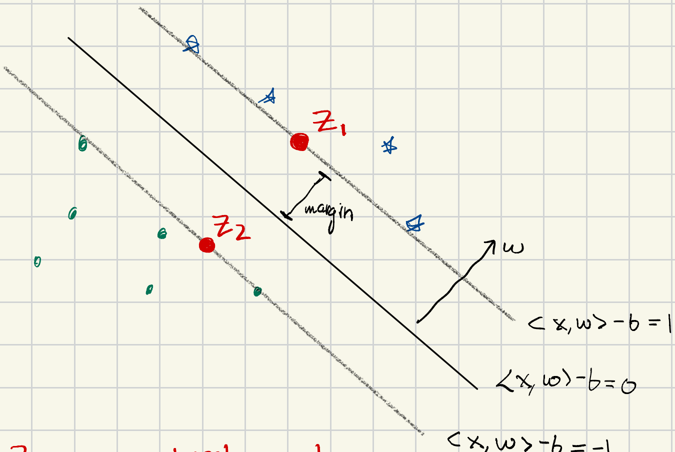
Q: Which separating line to choose?



A: Best margin



Computing the Margin



z_2 is closest point on "lower" hyper plane to z_1

$$\Rightarrow \quad z_1 - z_2 = \lambda \bar{w}$$

where $\bar{w} = \dfrac{w}{\|w\|_2}$

Q: what is $\|\bar{w}\|_2$?

$$\begin{aligned}
1 &= \langle z_1, w\rangle - b \\
&= \langle \lambda \bar{w} + z_2, w\rangle - b \\
&= \lambda \langle \bar{w}, w\rangle + \langle z_2, w\rangle - b \\
&= \lambda \frac{\langle w, w\rangle}{\|w\|_2} - 1 \\
&= \lambda \|w\|_2 - 1 \qquad \Leftrightarrow \lambda = \frac{2}{\|w\|_2}
\end{aligned}$$

$$\begin{aligned}
\text{Goal:} \quad & \max \quad \lambda \\
&= \max \quad \frac{2}{\|w\|_2} \\
&= \min \quad \frac{\|w\|_2}{2} \\
&= \min \quad \|w\|_2
\end{aligned}$$

Constrained Optimization

$$\min_{w,b} \|w\|_2 \quad \text{such that}$$

If $y^{(i)} = 1$, $\quad \langle x^{(i)}, w\rangle - b \geq 1$

If $y^{(i)} = -1$, $\quad \langle x^{(i)}, w\rangle - b \leq -1$

$$\Leftrightarrow \quad \min_{w,b} \|w\|_2 \quad \text{such that}$$

$$y^{(i)}\left[\langle x^{(i)}, w\rangle - b\right] \geq 1 \quad \forall i$$

9/25/25

Class Logistics

- Study sessions! Snacks!
- Tuesday

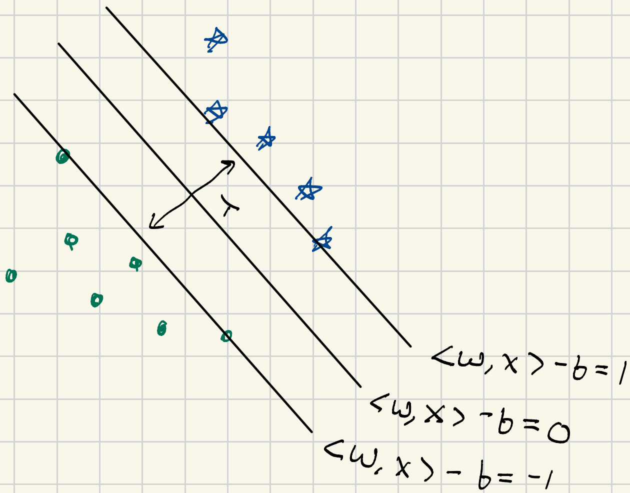
1	2	3	4	5
0	6	7	5	0

2.9

- Goal: challenging class with support
  - notes
  - discord
  - office hours
  - study sessions
- Pset due Tuesday (not Monday)
- Office Hours Next Week
  - Tuesday 12:30-2
  - Friday 10-11:30, 12:30-1:30
- No class Thursday, video instead

Review

Goal: Binary classification with largest margin



$$\arg\max_{w,b} \lambda \quad = \arg\min_{w,b} \frac{\|w\|_2}{2}$$

$$= \frac{2}{\|w\|_2}$$

$$\text{s.t.} \quad y^{(i)}\left[\langle w, x^{(i)} \rangle - b\right] \geq 1$$

If $y^{(i)} = 1$, then:

If $y^{(i)} = -1$, then:

Constrained optimization problem!

Today: How to solve ☺

Lagrangian

Goal: Reformulate as Unconstrained

$$\min_{w,b} \frac{\|w\|_2^2}{2} \quad \text{s.t.} \quad \underbrace{y^{(i)}\left[\langle w, x^{(i)}\rangle - b\right] \geq 1}_{y^{(i)}\left[\langle w, x^{(i)}\rangle - b\right] - 1 \geq 0} \quad \forall i$$

$$\Longleftrightarrow$$

Lagrangian $\mathcal{L}(w,b,\alpha)$

$$\min_{w,b} \max_{\alpha \in \mathbb{R}^n_+} \boxed{\frac{1}{2}\|w\|_2^2 - \underbrace{\sum_{i=1}^{n} \alpha_i \left[y^{(i)}\left[\langle w, x^{(i)}\rangle - b\right] - 1\right]}_{\text{If any constraint violated, this term gets arbitrarily large}}}$$

Lagrangian is convex ($\cup$) in w, b and concave ($\cap$) in α

Minimax Theorem:

$$\min_{w,b} \max_{\alpha} \mathcal{L}(w,b,\alpha) = \max_{\alpha} \min_{w,b} \mathcal{L}(w,b,\alpha)$$

Intuition: Saddle

Exact Optimization

$$\max_{\alpha} \min_{w,b} \mathcal{L}(w,b,\alpha)$$

Idea: Fix α, find optimal w, b

$$0 = \nabla_w \mathcal{L}(w^*, b^*, \alpha)$$

$$= w^* - \sum_{i=1}^{n} \alpha_i y^{(i)} x^{(i)} = 0$$

$$\Rightarrow w^* = \sum_{i=1}^{n} \alpha_i y^{(i)} x^{(i)}$$

$$0 = \frac{\partial}{\partial b} \mathcal{L}(w^*, b^*, \alpha)$$

$$= \sum_{i=1}^{n} \alpha_i y^{(i)} = 0$$

$$\frac{\partial}{\partial w_j} \frac{1}{2}\|w\|_2^2 = \frac{1}{2} \frac{\partial}{\partial w_j} \sum_{k=1}^{d} w_k^2$$

$$= w_j$$

$$\nabla_w \frac{1}{2}\|w\|_2^2 = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_d \end{bmatrix} = w$$

Show: $\nabla_w \langle x, w \rangle = x$

Returning to Lagrangian

$$\mathcal{L}(w,b,\alpha) = \frac{1}{2} w^T w - \sum_{i=1}^{n} \alpha_i \, y^{(i)} \langle w, x^{(i)} \rangle + \sum_{i=1}^{n} y^{(i)} \alpha_i b + \sum_{i=1}^{n} \alpha_i$$

$$w^* = \sum_{i=1}^{n} \alpha_i \, y^{(i)} x^{(i)}$$

$$\mathcal{L}(w^*, b^*, \alpha) = \frac{1}{2} \sum_{i=1}^{n} \alpha_i y^{(i)} x^{(i)T} \sum_{j=1}^{n} \alpha_j y^{(j)} x^{(j)} - \sum_{i=1}^{n} \alpha_i \, y^{(i)} \, x^{(i)T} \sum_{j=1}^{n} \alpha_j y^{(j)} x^{(j)} + \underbrace{b \sum_{i=1}^{n} y^{(i)} \alpha_i}_{=0} + \sum_{i=1}^{n} \alpha_i$$

$$= -\frac{1}{2} \sum_{i,j=1}^{n} \alpha_i \alpha_j \, y^{(i)} y^{(j)} \langle x^{(i)}, x^{(j)} \rangle + \sum_{i=1}^{n} \alpha_i$$

Dual Problem

$$\max_{\alpha} \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{n} \alpha_i \alpha_j y^{(i)} y^{(j)} \langle x^{(i)}, x^{(j)} \rangle \quad \text{s.t. } \alpha_i \geq 0 \text{ and } \sum_{i=1}^{n} \alpha_i y^{(i)} = 0$$

i : $\alpha_i > 0$ are the support vectors

How do we find w^*, b^* from solution α?

Why solve dual instead of orignal (primal) problem?

Soft Margin vs Hard Margin

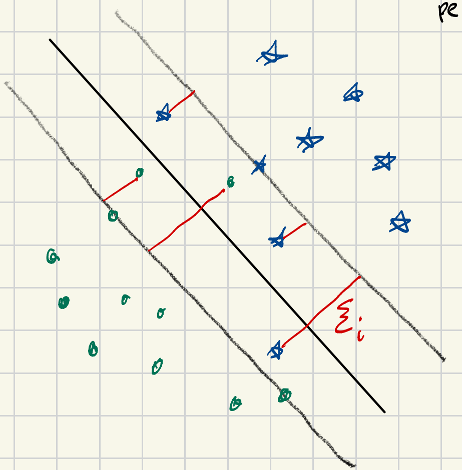
$\boxed{\text{Hard}}$ $\min_{w,b} \|w\|_2^2$ s.t. $y^{(i)}(\langle w, x^{(i)}\rangle - b) \geq 1 \quad \forall i$

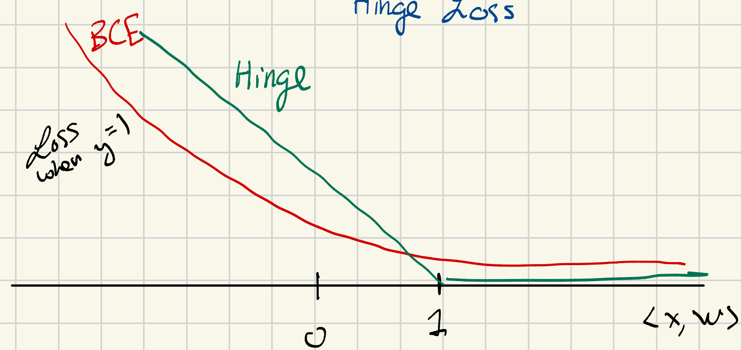
"slack variable"

$\boxed{\text{Soft}}$ $\min_{w,b} \|w\|_2^2 + C \sum_{i=1}^{n} \xi_i$ s.t. $\underbrace{y^{(i)}(\langle w, x^{(i)}\rangle - b) \geq 1 -}_{\xi_i \geq 1 - y^{(i)}(\langle w, x^{(i)}\rangle - b)} \xi_i, \quad \xi_i \geq 0 \quad \forall i$

penalty $C > 0$

$$\xi_i \overset{\text{best}}{=} \underbrace{\max(0, 1 - y^{(i)}(\langle w, x^{(i)}\rangle - b))}_{\text{Hinge Loss}}$$





Intuition: Hinge + regularization