support vector machine (SVM)

warm up: logistic regression

1		
$\vdots$		
i	x_i^T	y_i
$\vdots$		
n		

$s_i = x_i^T\beta$ ← logit score

$$p_i = \frac{e^{s_i}}{1+e^{s_i}}$$

$$1-p_i = \frac{1}{1+e^{s_i}}$$

$$s_i = \log\left(\frac{p_i}{1-p_i}\right) = \text{logit}(p_i) = \text{log odds ratio}$$

perceptron: $\hat{y}_i = \text{sign}(s_i)$ (hard decision)

logistic regression: soft perceptron

iterative reweighted LS

$$\beta_0 \xrightarrow{\Delta\beta} \beta_1 \longrightarrow \cdots \beta_t \xrightarrow{\Delta\beta} \beta_{t+1} \longrightarrow \cdots$$

$$\sum_{i=1}^{n} \hat{w}_i(\hat{y}_i - x_i^T\Delta\beta)^2 = \sum_{i=1}^{n}\left(\sqrt{\hat{w}_i}\,\hat{y}_i - \sqrt{\hat{w}_i}\,x_i^T\Delta\beta\right)^2$$

↓

$h_t(x_i)$ for xtboost

generalized linear model (GLM)

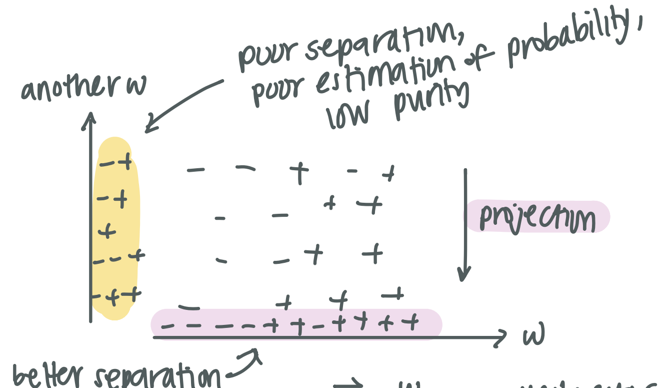
linear $x_i^T\beta$ + non-linear (e.g. sigmoid)

↓ learned ↓ specified

---

geometry $s_i = x_i^T w + b = \langle x_i, w\rangle + b$

w ↓ weight coefficient; b ↓ bias intercept



$\vec{u} = \frac{w}{|w|_{\ell 2}}$ unit vector

$\text{len}(\vec{u}) = 1$

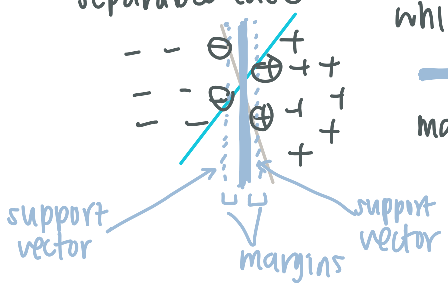
coordinate of projection:

$\langle x_i, \vec{u}\rangle$

regularization: $\frac{1}{2}|w|_{\ell 2}^2$ or $|w|_{\ell 1}$, we do not regularize b

geometry of SVM

separable case



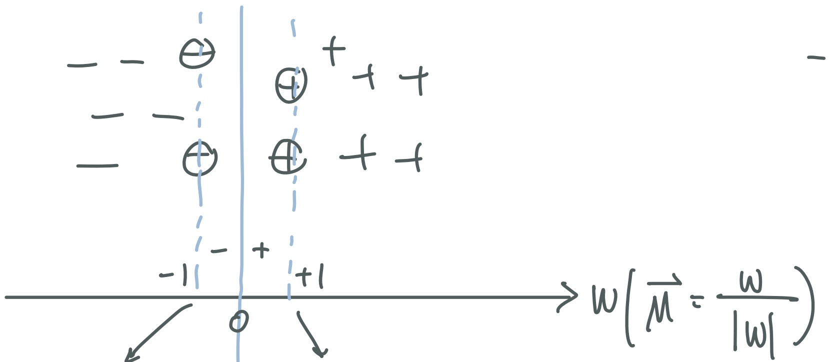
which plane is best among the 3?

gives you largest margin of error

max margin classifier

↓

margin of error



$\langle x_i, w\rangle + b = -1$ $\qquad$ $\langle x_i, w\rangle + b = +1$

$-1/+1$ arbitrary / by convention

0 represents boundary / separating hyperplane

positive support vectors x_i

$\langle x_i, w\rangle + b = +1 \rightarrow \langle x_i, \vec{\mu}\rangle = \frac{1-b}{|w|}$ $\quad$ coordinate for PSV after projection

For all positive examples

$\langle x_i, w\rangle + b \geq 1$ ①

negative support vectors x_i

$\langle x_i, w\rangle + b = -1 \rightarrow \langle x_i, \vec{\mu}\rangle = \frac{-1-b}{|w|}$

For all negative examples

$\langle x_i, w\rangle + b \leq -1$ ②

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

to maximize margin, we want to minimize w

$\langle x_i, \vec{\mu}\rangle = \langle x_i, \frac{w}{|w|}\rangle = \frac{\langle x_i, w\rangle}{|w|}$ $\quad$ coordinate for NSV after projection

max margin : primal problem

$$\boxed{\begin{array}{l} \min \frac{1}{2}|w|^2 \\ \text{subject to ① and ②} \\ y_i(\langle x_i, w\rangle + b) \geq 1 \\ i = 1, \ldots, n \end{array}}$$

convex constraint optimization problem.

NSV's : $\langle x_i, \tilde{w}\rangle + \tilde{b} = a_- \xrightarrow{\times c + d} -1$

PSV's : $\langle x_i, \tilde{w}\rangle + \tilde{b} = a_+ \xrightarrow{\times c + d} +1$

Lagrangian

$\alpha = (\alpha_1, \dots, \alpha_i, \dots \alpha_n)$ $\quad \alpha_i \geq 0$

$$L((w,b),\alpha) = \frac{1}{2}|w|^2 + \sum_{i=1}^{n} \alpha_i \left(1 - y_i(\langle x_i, w\rangle + b)\right)$$

constraint ≤ 0

$L((w,b)$ → primal parameters

α → dual parameters / Lagrangian multipliers

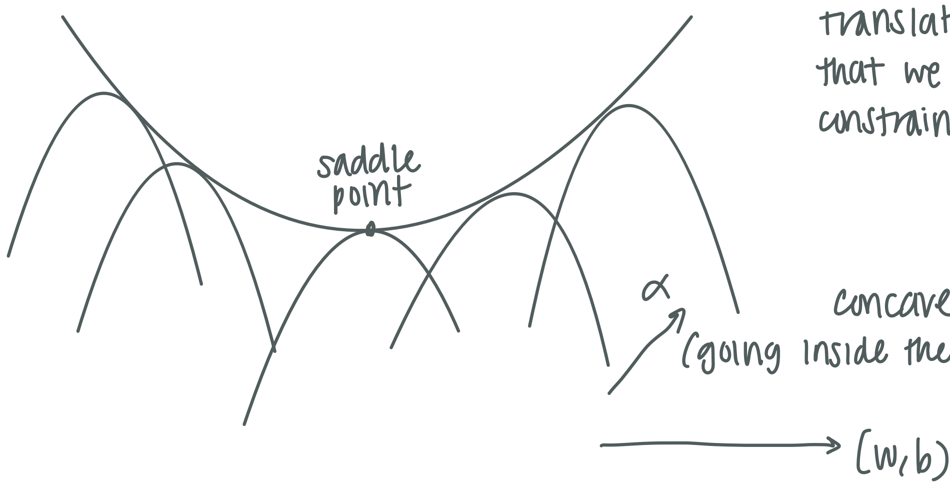
idea is to solve an unconstrained problem. b/c it is easier to solve an unconstrained problem.

Translation of primal problem to remove constraints

primal problem $\Longrightarrow \min_{(w,b)} \max_{\substack{\alpha : \alpha_i \geq 0 \\ i=1,\dots,n}} L((w,b), \alpha)$

remove original constraint by moving it into the objective function. (the Lagrangian)

translation into min/max problem so that we do not need to explicitly impose constraints



concave in α (going inside the blackboard/paper)

convex in (w,b)

why equivalence?

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

if there is i, $1 - y_i(\langle x_i, w\rangle + b) > 0$

there should not be any i where this constraint is violated.

$$\max_{\alpha_i > 0} \underbrace{\alpha_i}_{\infty} \underbrace{\left(1 - y_i(\langle x_i, w\rangle + b)\right)}_{>0} \longrightarrow \infty$$

primal problem $\Longrightarrow \min_{(w,b)} \max_{\substack{\alpha : \alpha_i \geq 0 \\ i=1,\dots,n}} L((w,b),\alpha) \xRightarrow[\text{convex}]{\text{concave-}} \max_{\alpha > 0} \boxed{\min_{(w,b)} L((w,b),\alpha)}$

solved in closed form $Q(\alpha)$

Dual problem: $\max_{\alpha > 0} Q(\alpha)$

duality: solution to 1 problem is the solution to another problem