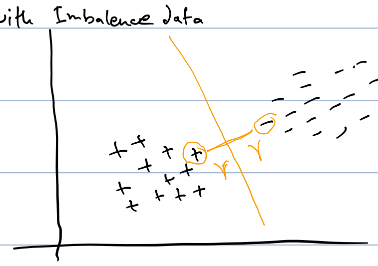
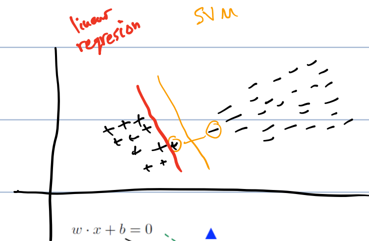
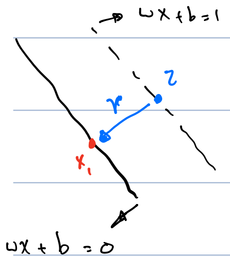
* another Linear classifier

doesn't have problem with Imbalence data



* deal with completly seperated classes
- we call it Maximum margin classifier
* $y = \{-1, +1\}$ * $y(wx + b) \geq 1$



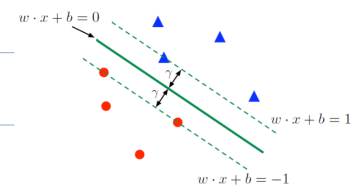


$$x_1 = z - \gamma \frac{w}{\|w\|}$$

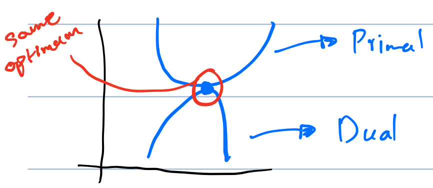
$$w\left(z - \gamma \frac{w}{\|w\|}\right) + b = 0$$

$$\frac{wz + b = 1}{-\gamma \frac{\|w\|^2}{\|w\|} = -1} \Rightarrow \gamma = \frac{1}{\|w\|}$$

we want large γ



So we need to minimize $\|w\|$ or $\frac{1}{2}\|w\|^2$ S.t. $y_i(wx_i + b) \geq 1$

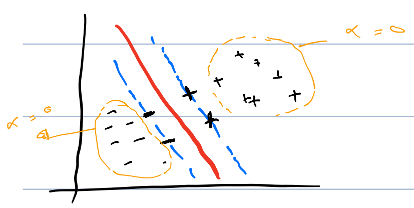


$\min \frac{1}{2}\|w\|^2$ S.t. $y^i(wx_i + b) \geq 1$ → Primal

$$\Downarrow$$

$$w = \sum \alpha_i y^{(i)} x^{(i)}$$
$$b = \sum \alpha_i y^{(i)}$$

$$\Downarrow$$



scaler $\alpha \geq 0$ $\{-1, 1\}$ vectors

$\max \sum \alpha_i - \frac{1}{2}\sum \alpha_i \alpha_j y^{(i)} y^{(j)} (x^{(i)} \cdot x^{(j)})$ → Dual

S.t. $\sum \alpha_i y_i = 0$

$y^i(w \cdot x^i + b) = 1$

Why we move to dual form?

① just depend on one dot product

② if $\alpha = 0$, the sample is not support vector
and you just have to store that samples

③ doesn't affected by feature expantion since it depends just on # of samples

it will reduce the variance

Soft margin SVM

Does have problem with Imbalance data

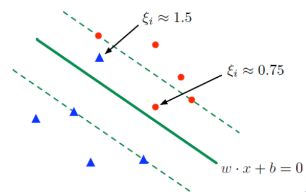
SV = on the margin and Vaulate sampler

→ allow some samples vaulate the condition

→ So the condition become $y^i(wx^i + b) \geq 1 - \xi_i$ slack $\xi_i \geq 0$

→ min $\frac{1}{2}\|w\|^2 + C \sum \xi_i$ hyperparameter

$\boxed{\sum \alpha_i y^i = 0}$



high C → hard margin SVM, but will solve when overlapping samples overfitting

low C → the margin will be as big as possible underfitting

Value of c will shift the decision boundry and change value of margin

$$\xi_i \begin{cases} 0, & y^i(wx^i+b) \geq 1 \quad \text{on or beyond the margin} \\ 1 - y^i(wx^i+b) & \text{inside the margin} \end{cases}$$

$\frac{1}{C}$ → So the affect is opposite

min $\underbrace{\sum \max(0, 1 - y(wx+b))}_{\text{Hinge loss}} + \lambda \frac{1}{2}\|w\|^2$ Primal

dual is same as hard svm with different condition

$\sum \alpha_i - \frac{1}{2} \sum\sum \alpha_i \alpha_j y^{(i)} y^{(j)} x^{(i)} \cdot x^{(j)}$ ← depends on # of samples (affected by # of features if you expanded)

$\|w\|^2 + C \sum \xi_i$ ← depends on # of features

$w = \sum \alpha y^i . x^i$

to classify → $\hat{y} = \text{Sign}\left(\sum \alpha y^i x^i . z + b\right)$ Still we have dot product between 2 expanded array of feature

Sample Vaulate $\alpha = C \implies y^i(wx + b) = 1 - \xi_i$

" on the margin $\alpha = (0, C) \implies y^i(wx + b) = 1$

❋ Kernal — to solve the product of 2 expanded

no need to expand then do dot product, do dot product before expanded

then apply the formula $(1 + x.z)^{p}$ p → degree

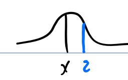
Kernal function ⇒ function that measure the similarity

you can replace dot product by your own Kernal function.

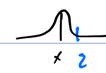
e.g. ① Jaccard Index $1 - \frac{x \cap z}{x \cup z}$

② Radial Basis Function $e^{-\frac{|x-z|^2}{\sigma^2}}$ → hyperparameter

high σ ⇒ underfitting



small σ ⇒ overfitting



more data ⇒ σ decreasing

overfit ! ⇒ increase σ

back to $\hat{y}$

to classify → $\hat{y} = \text{sign}\left[\sum (\alpha\, y^{i}\, x.z) + b\right]$

it's wieghted classifier

So it's like Knn, we measure How much the test sample similar to SVs

$$\underset{2\times 3}{\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}} \quad \underset{3\times 2}{\begin{bmatrix} 7 & 8 \\ 9 & 10 \\ 11 & 12 \end{bmatrix}}$$

$$\begin{array}{c|c} 7+18+33 & 8+20+36 \\ \hline 28+45+66 & 32+50+72 \end{array}$$