Support Vector Machines – Geometrical Interpretation of Max Margin

Zengchang Qin（Ph.D）

Support Vector Machine Interpretation

1) The Cost function of logistic regression

$$L(\theta) = \frac{1}{N}\sum_i y^{(i)}(-\log h_\theta(x^{(i)}) + (1-y^{(i)})(-\log h_\theta(x^{(i)}))$$

If we impose a regularization of $L(\theta)$

$$L(\theta) = \frac{1}{N}\sum_i y^{(i)}(-\log h_\theta(x^{(i)}) + (1-y^{(i)})(-\log h_\theta(x^{(i)})) + \frac{\lambda}{2N}\|\theta\|_2^2$$

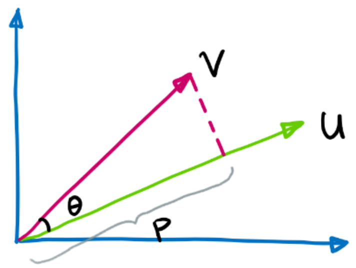
$$= C_1 \, Cost_1 + C_2 \, Cost_2 + \frac{1}{2}\|\theta\|_2^2 \qquad C \propto \frac{\lambda}{N}$$

$\min_\theta L(\theta) = C_1 \, Cost_1 + C_2 \, Cost_2 + \frac{1}{2}\|\theta\|_2^2$ is equivalent to

$$\boxed{\begin{aligned} &\min \frac{1}{2}\|\theta\|^2 \\ &\text{s.t.} \begin{cases} \theta^T x^{(i)} \geq 1 \\ \theta^T x^{(i)} \leq -1 \end{cases} \end{aligned}}$$

2) Dot product of two vectors: $u^T v = \|u\| \cdot \|v\| \cdot \cos\theta$

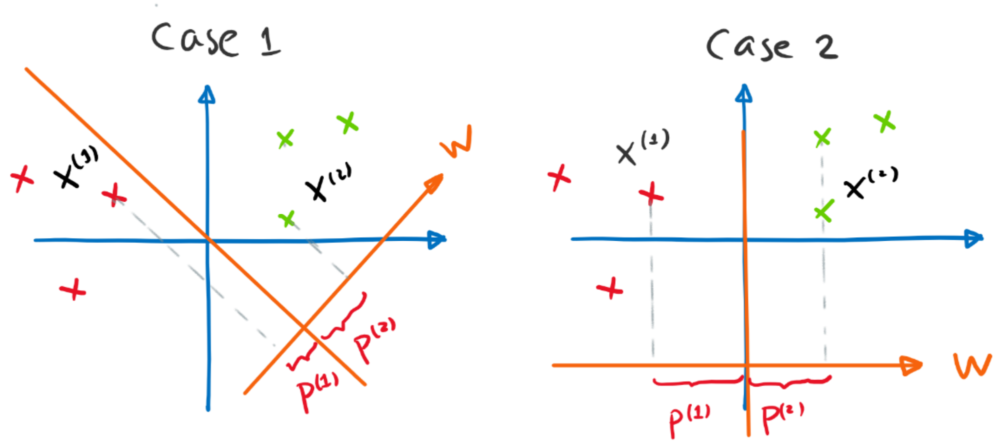
$$= P \cdot \|u\|$$



where p is the length of shadow of V on u.

The original optimization problem becomes:

$$\boxed{\begin{aligned} &\min \frac{1}{2}\|w\|^2 \\ &\text{s.t.: } p^{(i)}\|\theta\| \geq 1 \\ &\qquad\; p^{(i)}\|\theta\| \leq -1 \end{aligned}}$$



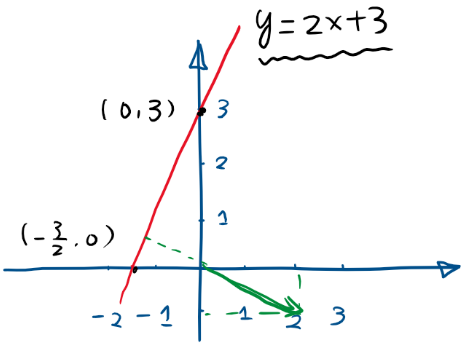
$p^{(i)}$ becomes bigger. $\|\theta\|$ becomes smaller

Case 2 is more preferred than the Case 1.

Case 2 corresponds to the max margin solution.

Appendix: Direction of W

1) Linear Function

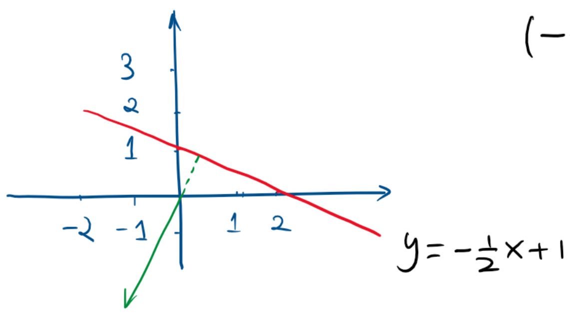


Given a linear function $y = 2x+3$, we have $(0,3)$ and $(-\frac{3}{2}, 0)$ on the hyperplane.

$$y=2x+3 \Rightarrow$$
$$2x-y+3=0$$
$$(2,-1)\begin{pmatrix}x\\y\end{pmatrix}+3=0$$

For a linear function $w^Tx+b=0$, the direction of w^T is perpendicular to the original linear function.

E.g.



$$(-\frac{1}{2},-1)\begin{pmatrix}x\\y\end{pmatrix}+1=0$$
$$w^T=(-\frac{1}{2},-1)$$