Recap: Dual form of SVM:

↳ $$\min_{\underline{\alpha}} \underbrace{\frac{1}{2} \alpha_n \alpha_m y_n y_m \underline{x}_n^T \underline{x}_n - \sum_{n=1}^{N} \alpha_n}_{-\mathcal{L}(\underline{\alpha})}$$

s.t. $\alpha_n \geq 0, \quad \forall n$

$\sum_{n=1}^{N} \alpha_n y_n = 0$

QP solver $\longrightarrow \underline{\alpha}^*$

↳ Then $\underline{w}^* = \sum_{n=1}^{N} \alpha_n^* y_n \underline{x}_n = \sum_{\{n;\, \alpha_n^* > 0\}} \alpha_n^* y_n \underline{x}_n$

$b^* = y_s - \sum_{\{n;\, \alpha_n^* > 0\}} y_n \alpha_n^* \underline{x}_n^T \underline{x}_s$, for some SV $\{\underline{x}_s, y_s\}$

↳ Decision rule:

$\hat{y} = \text{sign}(\underline{w}^{*T} \underline{x} + b^*)$

$\underline{w}^{*T} \underline{x} = \sum_n \alpha_n^* y_n \underline{x}_n^T \underline{x}$

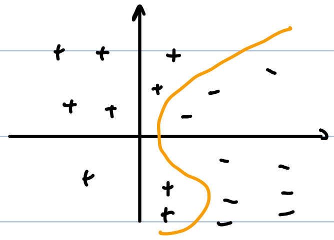
↳ key observation:

Dual problem and decision rule depends on $\{\underline{x}_1, \cdots \underline{x}_N\}$ only through $\underline{x}_n^T \underline{x}_m$

1. SVM and non-linear transform: (ch 8.3)

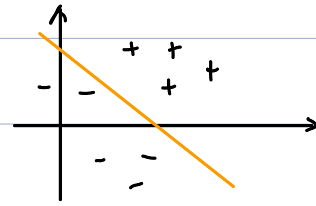
z-space

e.g.



$\underline{z} = (1, x_1, x_2, x_1^2, x_2^2, x_1 x_2)$ $\longrightarrow$

(quarditic transform)



↳ In z-space:

$$\mathcal{L}(\underline{\alpha}) = -\frac{1}{2} \sum_{n=1}^{N} \sum_{m=1}^{N} y_n y_m \alpha_n \alpha_m z_n^T z_m + \sum_{n=1}^{N} \alpha_n$$

Computing of QP: $Q \in \mathbb{R}^{N \times N}$

essentially unchanged!

Higher dimensional $\underline{z}$ ≡ better fit into dataset

SVM built-in defense against over-fitting.

↳ But is it costy to compute $\underline{z}_n^T \underline{z}_m$?

"kernel" $\triangleq k(\underline{x}_n, \underline{x}_m)$

⇒ Can we compute the kernel $k(\underline{x}_n, \underline{x}_m)$ w/out computing $\underline{z}_n$ or $\underline{z}_m$?

⇒ "kernel Method"

2. kernel Method:

① Polynomial kernel

e.g. $k(\underline{x}, \underline{x}') = (1 + \underline{x}^T \underline{x}')^2$ // $d = 2$, $\underline{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$

$= (1 + x_1 x_1' + x_2 x_2')^2$

$= 1 + x_1^2 x_1'^2 + x_2^2 x_2'^2 + 2x_1 x_1' + 2x_2 x_2' + 2 x_1 x_1' x_2 x_2'$

$= \underline{z}^T \underline{z}'$ if

$\underline{z} = (1, x_1^2, x_2^2, \sqrt{2}x_1, \sqrt{2}x_2, \sqrt{2}x_1x_2)$

$\underline{z}' = (1, x_1'^2, x_2'^2, \sqrt{2}x_1', \sqrt{2}x_2', \sqrt{2}x_1'x_2')$

Identical to qudratic transform w/ properly re-scaled $\underline{w}$.

↳ <u>General polynimal kernel:</u>

Use $k(\underline{x}, \underline{x}') = (1 + \underline{x}^T\underline{x}')^Q$ // Q means integer

$= (1 + x_1x_1' + \cdots + x_dx_d')^Q$ // Q th polynomial

= inner product of Q-th order polynomial

By running SVM:

$$\mathcal{L}(\underline{\alpha}) = -\frac{1}{2}\sum_{n=1}^{N}\sum_{m=1}^{N} y_n y_m \alpha_n \alpha_m k(\underline{x}_n, \underline{x}_m) + \sum_{n=1}^{N}\alpha_n$$

$$\hat{y} = \text{sign}\left(\sum_n \alpha_n y_n k(\underline{x}_n, \underline{x}) + b\right)$$

$$b = y_s - \sum_n \alpha_n y_n k(\underline{x}_n, \underline{x}_s)$$

② Fancier kernel:

$k(\underline{x}, \underline{x}') = e^{-r(\|\underline{x} - \underline{x}'\|^2)}$, for some const $r > 0$

<u>Fact:</u> z-space is infinite dimensional

e.g. $r=1$, $d=1$

$$k(x, x') = e^{-(x-x')^2} = e^{-x^2}e^{-x'^2}e^{2xx'}$$

$$= e^{-x^2}e^{-x'^2}\sum_{k=0}^{\infty}\frac{2^k x^k x'^k}{k!}$$ ← Taylor expansion

$$\therefore z = e^{-x^2}\left[1, \frac{\sqrt{2}x}{\sqrt{1!}}, \frac{\sqrt{2^2}x^2}{\sqrt{2!}}, \frac{\sqrt{2^3}x^3}{\sqrt{3!}} \cdots\right]$$

$$z' = e^{-x'^2}\left[1, \frac{\sqrt{2}x'}{\sqrt{1!}}, \cdots\right]$$