Aprendizaje Estadistico

Introduccion:

En Aprendizaje estadistico queremos aprender sobre datos.

Generalmente tenemos:

- Un resultado (quantitativo: precio de acciones
  Categorico: cancer esta presente o no)

- datos
- Set de entrenamiento
- Modelo de prediccion.

Que es lo que queremos hacer?

- Predecir casos que no hemos visto.
- Como nuestros datos affectan la prediccion
- Queremos ver la calidad de predicciones.

K vecinos mas cercanos

K=1, no fue a la emergencia.

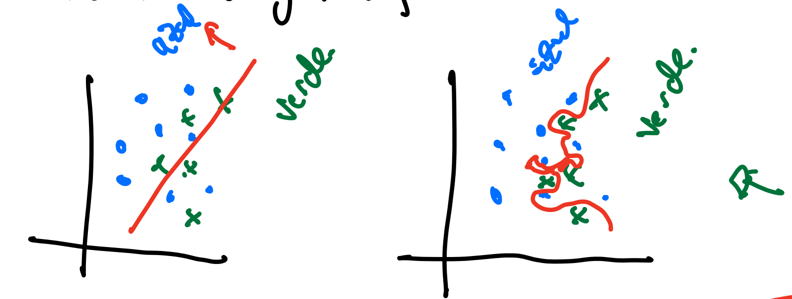
K=3, no fue a la emergencia.

$$\hat{Y}(x) = \frac{1}{K} \sum_{x_i \in N_K(x)} y_i$$

Donde $N_K(x)$ es el vecindario de x definido por los K puntos x_i en el set de entrenamiento.

Lo que queremos es una curva de decisión.

- Cuando K es pequeña, la curva cambia mucho.
- Cuando K es grande, la curva no cambia mucho (es suave).



K $\boxed{\text{grande}}$ K $\boxed{\text{pequeña}}$

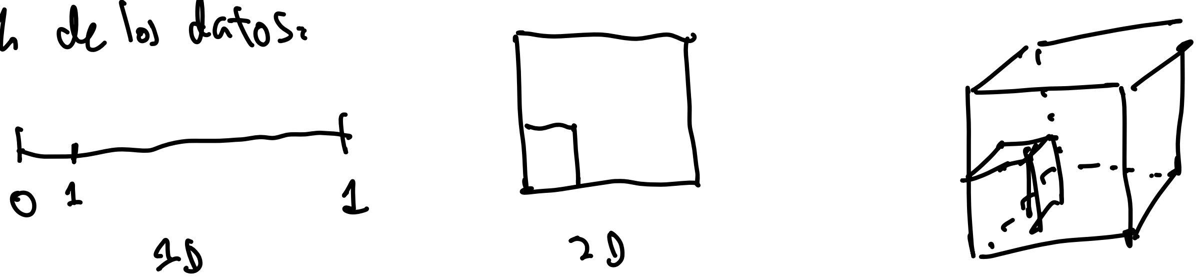
grande 10
pequeña 2

grande 4
pequeña 20

Metodos locales en altas dimensiones

"La Maldición de la Dimensionalidad"

Distancia entre puntos puestos uniformes si queremos capturar el 10% de los datos.



1D: $\ell = 0.1$

2D: $\ell^2 = 0.1 \Rightarrow \ell = 0.32$

3D: $\ell^3 = 0.1$ $\ell = 0.46$

Muy larga la distancia.

Teoría De decisión estadística

$Y = f(X)$, queremos encontrar la f que minimiza una función de pérdida.

Por ejemplo: $L(Y, f(x)) = (Y - f(x))^2$.

Estamos interesados del error de predicción (EPE).

$$EPE(f) = E\left[(Y - f(x))^2\right]$$

$$= \int [y - f(x)]^2 \Pr(x, y)\, dx\, dy.$$

Si condicionas sobre x (un dato que tenemos)

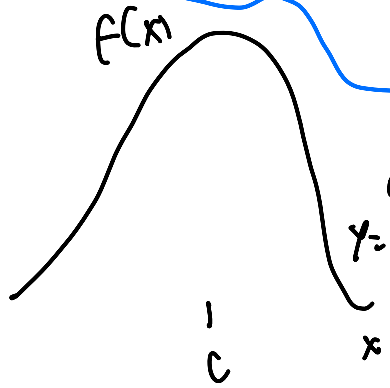
$$EPE(f) = E_x\left[E_{Y|x}\left((Y - f(x))^2 \mid x\right)\right]$$

Lo que queremos hacer es minimizar el EPE para cualquier valor de x:

$$f(x) = \underset{c}{\operatorname{argmin}} \left[E_{Y|x}\left([Y - c]^2 \mid X = x\right)\right]$$

La solución es

$$f(x) = \boxed{E(Y \mid X = x)}$$



$$y = \underset{c}{\operatorname{argmin}}\, E\left[(x - c)^2\right]$$

$$Y = \boxed{E(x)} \;?$$

— Día 2 — Se va a transmitir por TV UNAH

$\left[Y = f(x) + \epsilon, \quad E(\epsilon) = 0, \quad \text{Varianza}(\epsilon) = \sigma^2 \right.$

$$EPE_k(x_0) = E\left[(Y - \hat{f}_k(x_0))^2 \mid X = x_0\right] \ldots$$

$$= \sigma^2 + \text{Sesgo}^2(\hat{f}_k(x_0)) + \text{Var}(\hat{f}_k(x_0))$$

error irreducible

$$= \sigma^2 + \left[f(x_0) - \frac{1}{K}\sum_{\ell=1}^{K} f(x_\ell) \right]^2 + \frac{\sigma^2}{K}$$

Para un $f(x)$, cuando K aumenta, [Para un $f(x)$ suave]

$\left[f(x_0) - \frac{1}{K}\sum_{\ell=1}^{K} f(x_\ell) \right]^2$ $\boxed{\text{Aumenta}}$ ↗ aumenta $\boxed{20}$ ↘ disminuye $\boxed{7}$

$\frac{\sigma^2}{K}$ $\boxed{\text{disminuye}}$ ↗ aumenta $\boxed{}$ ↘ disminuye $\boxed{}$ todos

Compromiso del cesgo y la varianza.