$$MSE = \frac{1}{2}\sum_{i=1}\left(y_i - \hat{f}(x_i)\right)^2$$

MSE se puede tomar p/la data de training pero lo mas interesante es medirlo en la data de test para ver como performa con data no vista antes.

Entonces podemos separar
Train MSE

$$E\left(Y_0 - \hat{f}(x_0)\right)^2 = Var\left(\hat{f}(x_0)\right) + \left[Bias\left(\hat{f}(x_0)\right)\right]^2 + Var(\varepsilon)$$

Y_0, x_0 son los valores de test.

Var() es cuanto estimamos que cambia $\hat{f}$ cuando la calculamos con otro training set.
Idealmente $\hat{f}$ no debería cambiar mucho si la entrenamos con otro <u>training</u> <u>set</u>.

- Modelos mas flexibles tienen varianza mas alta, esto es porque pequeños cambios en el training set generan grandes cambios en $\hat{f}$

- Cuando la curva está muy cerca de los puntos un cambio en un punto hace que $\hat{f}$ tenga que cambiar mucho.

Bias: El error que se introduce al aproximar $\hat{f}$ a un problema real ($f()$)

+ Flexibilidad => ↑ VAR ↓ Bias.
Los cambios relativos entre VAR y Bias determinan si MSE aumenta o disminuye

Clasificación

accuracy $\hat{f}$ Proporción de errores al estimar

$$\frac{1}{n} \hat{\sum} I(y_i \neq \hat{y}_i)$$

$$\begin{aligned} I(y_i \neq \hat{y}_i) &= 1 \quad \text{si } y_i \neq \hat{y}_i \\ &= 0 \quad \text{si } y_i = \hat{y}_i \end{aligned}$$

ERROR rate

Regresión lineal (supervised learning)

$Y \approx \beta_0 + \beta_1 X$

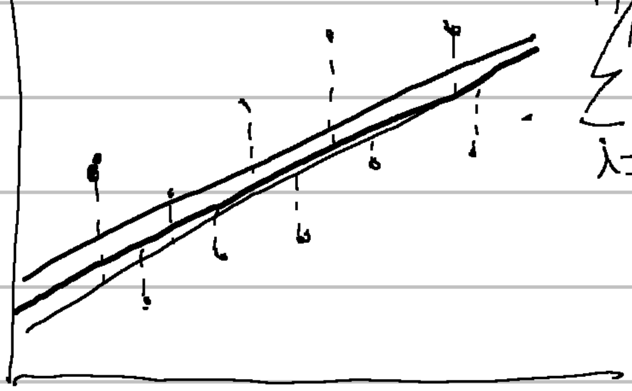
$\downarrow$

modelado aproximadamente por.

$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 x$

parámetros a determinar.

Y $y_1 \; y_2 \; \dots \; y_n$ (y_i).



$$\sum_{i=1}^{n} \left(y_i - \hat{y}_i\right)^2 = \min$$

X. $x_1, x_2 \dots x_n$ (x_i)

$y_i = m x_i + b.$

$(y_i, x_i).$

$y = \overset{\text{pendiente}}{m} x + b$ ordenada al origen

Minimizar: Cuadrados mínimos

$e_i = y_i - \hat{y}_i$ (residuo i)

$RSS = \sum e_i^2$

$\uparrow$

función a minimizar.

Minimizando

$\hat{\beta}_1 = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}$ $\beta_0 = \bar{y} - \hat{\beta}_1 \bar{x}$ (Demostrar)

$\bar{y}, \bar{x}$: promedios de los muestras

Resampling → Cross Validation
Bootstrap

CV

Puede usarse p/ estimar el test error.

Test-error-rate : errores al usar el modelo en data nunca vista

Training-error-rate: el error que se obtiene al correr el modelo contra la data que se testeo.

Con CV se estima el test-error con un subset del training set.

Leave-One-Out Cross Validation (LOOCV)

K-Fold

Se divide el training set en K partes.
Se toma el primer fold y luego se usan las restantes para entrenar.

Se repite tomando el segundo fold como validación y así hasta el último.

Se tienen los K MSE y se promedian.

4) Clasificación

Probabilidad que una observación pertenezca a una clase.

Métodos

- regresión logística
- naive Bayes
- KNN

- ✓ linear discriminant analysis
- ✓ quadratic " "
- ✓ Poisson regression

La clasificación que se obtiene usando una regresión lineal para predecir una respuesta binaria es la misma que si usamos "linear discriminant analysis" (✓).

La regresión logística modela la probabilidad de pertenecer a una clase determinada

Regresión logística

$$\boxed{P(Y = True | X)} \qquad p(X)$$

$Y = \text{True}$ si $p(x) > 0,5$ -

$p(x) > 0,1$ conservador.

Modelo

$p(x) = \beta_0 + \beta_1 X.$

$p(X) = P_r(Y = 1 | X)$

Si usamos la regresión lineal tenemos el problema de que hay valores menores a cero y mayores a 1 que no tiene sentido modelarlo como probabilidad.

$\Rightarrow$ Modelamos la regresión p/ que devuelva valores entre 0 y 1

$$p(X) = \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}} \qquad 1 - p(x) = 1 - \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}}$$

$$\Rightarrow \frac{1 + e^{\beta_0 + \beta x} - e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}} = \frac{1}{1 + e^{\beta_0 + \beta_1 x}}$$

$$\Rightarrow \text{odd} = \frac{p(x)}{1 - p(x)} = \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}} \cdot 1 + e^{\beta_0 + \beta_1 x} = e^{\beta_0 + \beta_1 x}$$

$$\therefore \boxed{\underline{\text{odds}} = \frac{p(x)}{1-p(x)} = e^{\beta_0 + \beta_1 x}}$$

odds $\cong 0$ $\Rightarrow$ baja probabilidad } de $p(x) = 1$
odds $\cong \infty$ $\Rightarrow$ Alta " }

Ejemplo 1

1/5 de las personas con odds de 1/4 defaultean.

$$p(x) = 1/5 = 0{,}2 \Rightarrow \frac{p(x)}{1-p(x)} = \frac{0{,}2}{1-0{,}2} = \frac{1/5}{4/5} = \frac{1}{8}\,\frac{8}{4} = \boxed{\frac{1}{4}}$$

Ejemplo 2

9 de cada 10 con odds de 9 defaultean.

$$p(x) = 9/10 \Rightarrow \frac{9/10}{1/10} = \frac{9}{\cancel{10}}\,\frac{\cancel{10}}{1} = \boxed{9 = \text{odds}}$$

$$\boxed{p(x) = k \Rightarrow \text{odds} = \frac{k}{1-k}}$$

Recordemos

$$\frac{p(x)}{1-p(x)} = \text{odds} = e^{\beta_0+\beta_1 x}.$$

$$\Rightarrow \boxed{\ln\left(\frac{p(x)}{1-p(x)}\right) = \beta_0 + \beta_1 x}$$

log-odd or logit $= \beta_0 + \beta_1 X.$

Para la regresión lineal $Y = \beta_0 + \beta_1 x$

β_1 da el cambio promedio de y cuando X cambia en una unidad.

En la logística aumentar X en una unidad cambia log(odd) en β_1