Lect 20 4/25/18

Previously in Binary Classification, we precised error via:

$$MAE = \frac{1}{n}\sum \mathbb{1}_{y_i \neq \hat{y}_i} \quad (\text{misclassification error}/\text{rank})$$

There are two types of errors:

* $y_i = 0$ & $\hat{y}_i = 1$ (called false positive)
* $y_i = 1$ & $\hat{y}_i = 0$ (called false omission).

How can you visualize these type errors?

Predict° $(\hat{y})$

FP = False Positive
FN = " Neg
TN = True Neg.
TP = True Posi

		Predict 0	Predict 1	
true (y)	0	TN	FP	#N
	1	FN	TP	#P
		#PN	#PP	n

#PN : Nber of predictive Neg
#PP : Nber of " Posi

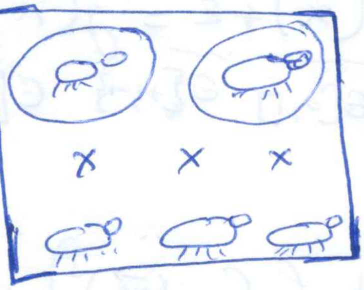
Called CONFUSION TABLE

$$\boxed{\text{Misclassification error} = \frac{FP + FN}{n}}$$ or $\text{Accuracy} = 1 - \frac{FP + FN}{n}$

or $= \frac{TP + TN}{n}$

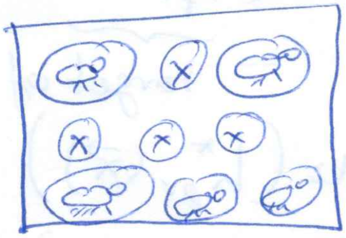
$$\text{Precision} = \frac{TP}{\#PP}$$

$$\text{Recall or Sensitivity} = \frac{TP}{\#P}$$



precision $= \frac{2}{2} = 100\%$

recall $= \frac{2}{5} = 40\%$



Precision $= \frac{5}{9} = 56\%$

Recall $= \frac{5}{5} = 100\%$

$$F_1 := \frac{2}{\frac{1}{\text{Recall}} + \frac{1}{\text{Precis.}}}$$

F_1 x $> 2\%$

False discovery rate $= 1 - \text{precision} = \frac{FP}{\#PP}$
(FDR)

False omission rate $= \frac{FN}{\#PN}$

$y = [0,1] \leftarrow$

TN	FP
FN	TP

New Unit:

$y = \{0,1\}$

Targets is $P(Y=1|\vec{x}) \in [0,1]$

Prob. Estimate

Probability Classification.

$$Y = t(\bar{z}) = f(\vec{x}) + \delta = \underbrace{h^{*}(\bar{x})}_{\in[0,1]} + \underbrace{\varepsilon}_{\in\{+1,-1\}} = g(\bar{x}) + e \Rightarrow f, h^{*}, g, \text{do}$$

$t(\bar{z}) \in [0,1]$, $f(\vec{x}) \in [0,1]$, $\delta \in [+1,-1]$, $g(\bar{x}) \in [0,1]$, $e \in \{+1,-1\}$

not resu... prob's

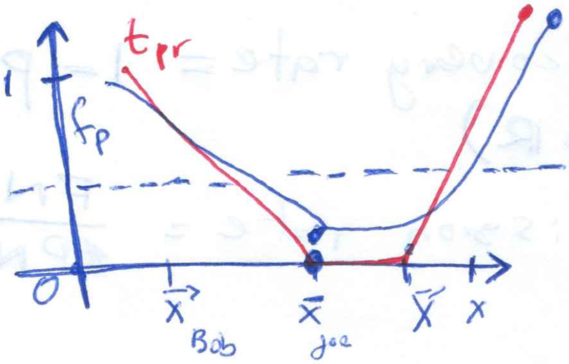
Consider: $Y \sim \text{Bernoulli}(\underbrace{f_{pr}(\vec{x})}_{\text{target of estimate}})$

$Y \sim \text{Bernoulli}(h^{*}_{pr}(\vec{x}))$

$Y \sim \text{Bernoulli}(g_{pr}(\vec{x}))$

$t(\bar{z}) = 1 \qquad g_{pr}(\vec{x}) = 0.2$

$t_{pr}, f_{pr}, h^{*}_{pr}, g_{pr}$



$$\boxed{\begin{array}{l} g_{pr} = \mathcal{A}(\mathcal{H}, \mathbb{D}) \\ \text{What is } \mathcal{H}, \mathcal{A}\text{ ?} \end{array}}$$

observe that:

$Y_1 \sim \text{Bern}(f_{p}(\vec{x}_1)) = f_{pr}(\vec{x}_1)^{Y_r}(1 - f_{pr})^{1-y_r}$

$Y_2 \sim \text{Bern}(f_{pr}, (\vec{x}_n))$

$\vdots$

$Y_n \sim \text{Bern}(f_{pr}, (\vec{x}))$

What is $p(Y_1, Y_2, \ldots, Y_n)$?

Unknown ... unless you know / assump

Big Assumpt: $Y_1, \dots, Y_n$ are Indpdt.

How to check this assumpt: Ask how $\mathbb{D}$ was collected

▶ Indepdt: If indpdt

$$\underbrace{P(Y_1, Y_2, \dots, Y_n | x)}_{\text{likelihood}} = \prod_{i=1}^{n} P(Y_i / \vec{x}_i) = \prod_{i=1}^{n} f_p(\vec{x})^{y_i} (1 - f_{pr}(\vec{x}))^{1-y_i}$$

We want to pick f_{pr} such that:

$P(Y_1, \dots, Y_r / X)$ is MAX.

But f_{pr} is arbitrary compleKated so we construct our candidate set to $\mathcal{H}$.

How about ... $\mathcal{H} = \{ \cancel{\phi} \vec{w} \cdot \vec{x} : \vec{w} \in \mathbb{R}^{p+1} \}$?

Illegal since $\vec{w} \cdot \vec{x} \in \mathbb{R}$ and $\forall\, b(\vec{x}) \leftarrow (0,1)$.

How about $\mathcal{H} = \{ \mathbb{1}_{\vec{w} \cdot \vec{x} \geq 0} : \vec{w} \in \mathbb{R}^{p+1} \}$? Illegal since $\mathbb{1}_{\vec{w} \cdot \vec{x}} \in \{0,1\}$ and we said we meet prob!

But, .. I still like $\vec{w} \cdot \vec{x}$! How can I use it?

We need p "link fun", ϕ, such that $\phi(\vec{w}, \vec{x}) \in (0,$

Because we can't be too sure and say prob = 0 or 1.

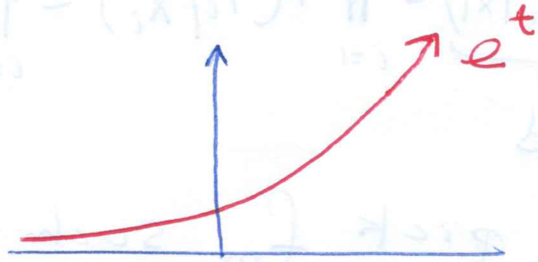
$\phi : \mathbb{R} \longrightarrow (0, 1)$ and is strictly increasing.

There are many ~~possibilities~~ possible ϕ's. Popular

Popular choices includ:

a) $\phi(x) = \underbrace{\frac{e^{n}}{1+e^{n}}}_{\text{(sigma)}} = \frac{1}{1+e^{-n}}$

b) $\phi(x) = \underbrace{\Phi^{-1}(c_i)}_{\text{(probit)}}$



c) $\phi(u) = 1 - e^{-u}$

This is called complementary log·log

d) Hyperbolic tangent:

$$\phi(u) = \tanh(u) = \left(\frac{e^{u} - e^{-u}}{e^{u} + e^{-u}}\right) \cdot \frac{1}{2}$$

▷ Logistic Regression

$$\mathcal{H} = \left\{ \frac{e^{\vec{w}\cdot\vec{x}}}{1+e^{\vec{w}\cdot\vec{x}}} : \vec{w} \in \mathbb{R}^{p+1} \right\}$$

Called Generalized Linear Model (GLM)

$$\text{Let } \vec{b} = \operatorname{argmax}\left\{ \prod^{n} \left(\underbrace{\frac{e^{\vec{w}\cdot\vec{x}}}{1+e^{\vec{w}\cdot\vec{x}}}}_{= \frac{1}{1+e^{-\vec{w}\cdot\vec{x}_i}} = \left(1+e^{-\vec{w}\cdot\vec{x}}\right)^{-1}} \right)^{y_i} \cdot \left(\underbrace{1 - \frac{e^{\vec{w}\cdot\vec{x}}}{1+e^{\vec{w}\cdot\vec{x}}}}_{= \frac{1}{1+e^{\vec{w}\cdot\vec{x}_i}} = \left(1+e^{\vec{w}\cdot\vec{x}_i}\right)^{-1}} \right)^{1-y_i} \right\}$$

$$\Rightarrow \vec{b} = \operatorname{argmax}\left\{ \prod_{i=1}^{n} \left(\left(1+e^{-\vec{w}\cdot\vec{x}}\right)^{-1} \right)^{y_i} \left(\left(1+e^{\vec{w}\cdot\vec{x}}\right)^{-1} \right)^{1-y_i} \right.$$

$$= \begin{cases} \left(1+e^{-\vec{w}\cdot\vec{x}}\right)^{-1} & \text{if } y_i = 1 \\ \left(1+e^{\vec{w}\cdot\vec{x}}\right)^{-1} & \text{if } y_i = 0 \end{cases}$$

$$= \prod_{i=1}^{n} \left(1+e^{(1-2y_i)\vec{w}\cdot\vec{x}_i} \right)^{-1}$$

$$= \prod^{n} \left(1+e^{-z_i\vec{w}\cdot\vec{x}_i} \right)^{-1}$$

$$\vec{b} = \left\{ -\sum \ln\left(1+e^{-z_i\vec{w}\cdot\vec{x}_i}\right) \right.$$

$$= \underset{\vec{w}}{\operatorname{argmin}} \left\{ \sum \ln\left(1+e^{-z_i\vec{w}\cdot\vec{x}_i}\right) \right\}$$

Take $\frac{d}{d\vec{x}}\left[\quad\right] \overset{\text{set}}{=} 0$ but no closed form analytical solut

Need: numerical optimizat:

Typically " Grad.