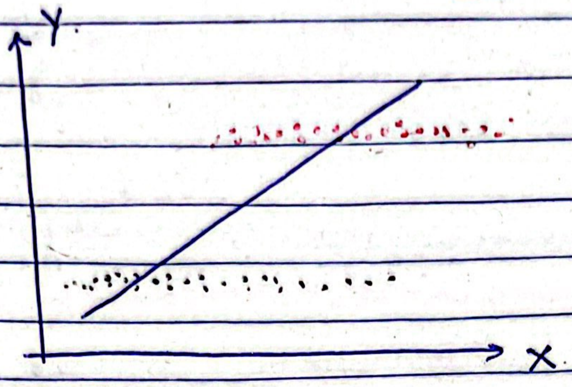
* In classification problems, the response $\boxed{y}$ is $\boxed{\text{categorical variable}}$

Why not linear regression?



Propensity score:

Suppose we consider y_i as a binary category:

$$y_i = \begin{cases} 1 & ; \text{ if } i\text{-th observation is in class 1.} \\ 0 & ; \text{ if } i\text{-th observation is in class 2.} \end{cases}$$

* We are not modelling the outcome directly.

* Consider the conditional probability, say $p(y_i = 1 \mid x_i)$.

$P(y_i = 1 \mid x_i)$ = The propensity score of class 1.

$P(y_i = 0 \mid x_i)$ = The propensity score of class 2.

$$P(y_i = 0 \mid x_i) = 1 - P(y_i = 1 \mid x_i).$$

Scanned with CamScanner

* Odds of an event.

$$\text{odds} = \frac{p}{1-p} = \frac{\text{probability that the event will occur}}{\text{probability that the event will not occur.}}$$

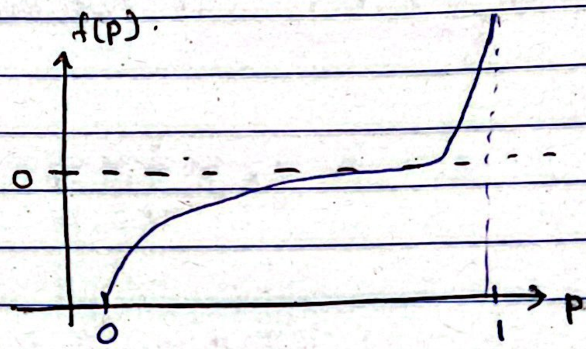
$$\text{Odds of being in class 1} = \underbrace{\frac{P(y_i = 1 \mid x_i)}{1 - P(y_i = 1 \mid x_i)}}_{\text{The ratio of the propensity scores of the two classes.}}$$

* Logit function.

$$f(p) = \log_e \left(\frac{p}{1-p}\right).$$

$$-\infty < f(p) < \infty$$

for all $p \in (0, 1)$.



* Logistic regression for binary response.

* Logistic regression is a generalized linear model. where it models the log odds as a linear combination of predictors:

$$\text{logit}(p_i) = \log_e\left(\frac{p_i}{1-p_i}\right) = \sum_{j=0}^{p} \beta_j x_{ij},$$

$$p_i = \frac{e^{\sum_{j=0}^{p} \beta_j x_{ij}}}{1 + e^{\sum_{j=0}^{p} \beta_j x_{ij}}}$$

* $y_i \sim Bin(1, p_i)$ where $p_i = P(y_i = 1 | x_i)$.

* The link function is logit function.

In GLM, the link function links the predictors to the model parameters.

* Logistic regression is a linear classifier.

* Threshold.

* We choose the threshold q such that $P(y_i = 1 | x_i) \geq q$ is considered to be in class 1.

* decision boundary : $\log\left(\frac{p}{1-p}\right) = \sum_{j=0}^{p} \beta_j x_{ij}$

eg:- If $p = 2$;

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2}.$$

$$x_{i2} = \underbrace{\frac{1}{\beta_2}\left[\log\left(\frac{p}{1-p}\right) - \beta_0\right]}_{\text{intercept}} - \underbrace{\frac{\beta_1}{\beta_2}}_{\text{slope}} x_{i1}.$$

Scanned with CamScanner

* Confusion matrix / Classification table.

	Actual condition.	
Prediction condition.	True Positive (TP)	False positive (FP). Type I error.
	False Negative (FN) Type II error	True Negative (TN).

True positive rate (TPR) = $\frac{TP}{TP + FN}$.

False positive rate (FPR) = $\frac{FP}{FP + TN}$.

Sensitivity = $\frac{TP}{TP+FN}$.

Specificity = $\frac{TN}{FP+TN}$.

Recall. = $\frac{TP}{TP+FN}$.

Precision = $\frac{TP}{TP+FP}$.

False discovery rate = $\frac{FP}{TP+FP}$.

Total population (n) = TP + FN + FP + TN.

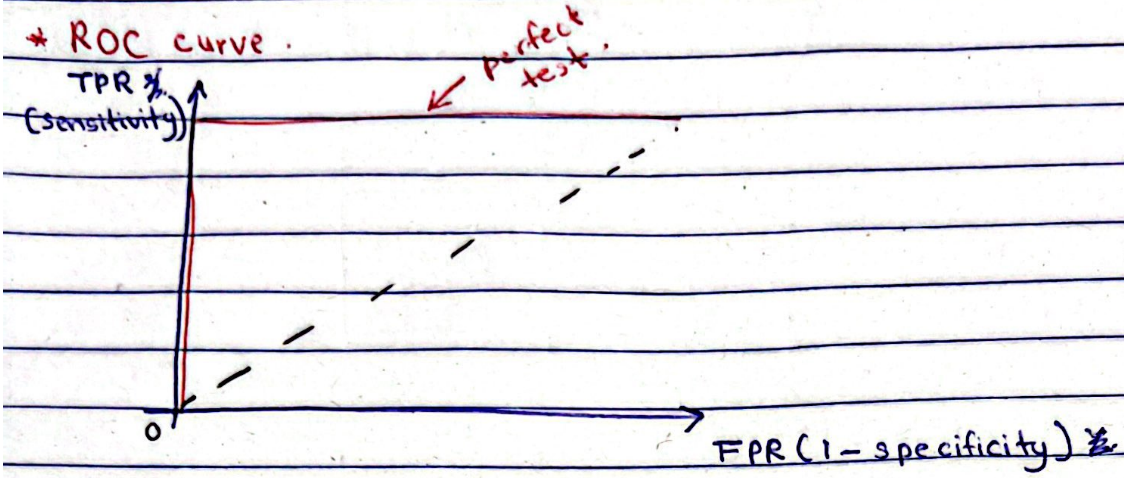
Prevalence = $\frac{TP+FN}{n}$

Accuracy = $\frac{TP+TN}{n}$.

Balanced accuracy = $\frac{TPR + FPR}{2}$.

F1 = $\frac{2 (\text{Precision} \times \text{Recall})}{(\text{Precision} + \text{Recall})}$.

* ROC curve.



⟹ If AUC is lower, it's best.