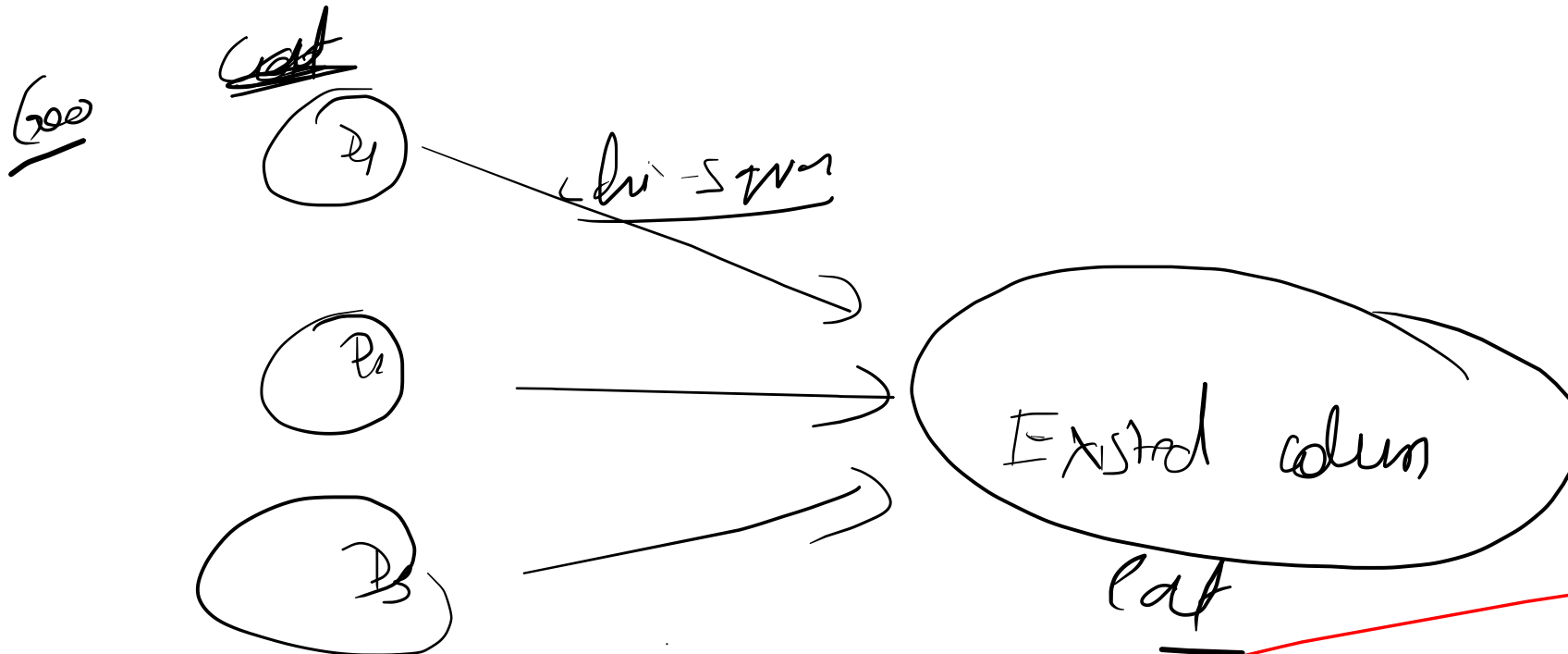
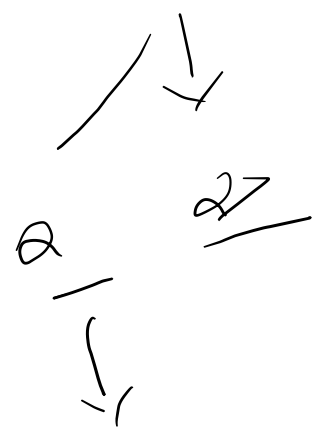
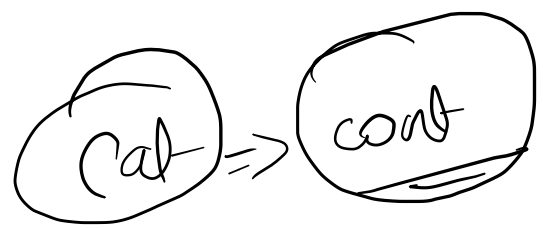


churn modelling





hard encoding

label encoding

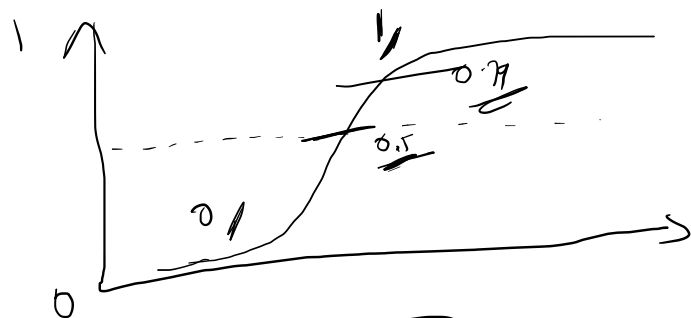
M	0
M	0
F	1
M	0
F	0
M	0
F	1
F	0

	<u>SUV</u>	<u>LUV</u>	<u>MUV</u>	X
✓ <u>SUV</u>	1	0	0	
✓ <u>LUV</u>	0	1	0	
✓ <u>MUV</u>	0	0	1	
SUV	1	0	0	
LUV	0	1	0	
MUV	0	0	1	



One hot encoding

Label binarizer
or
dummy variable



$$p = \frac{1}{1 + e^{-(\theta_0 + \eta x)}}$$

0.5

FPR

actual

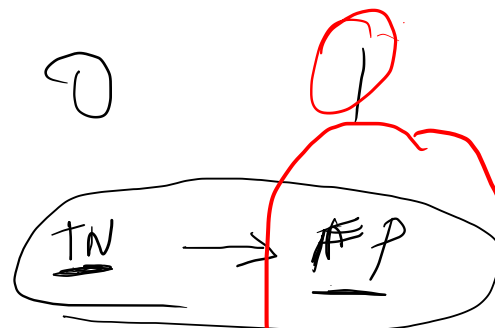
TPR

$$\text{Accuracy} = \frac{TN + TP}{TN + TP + FP + FN}$$

$$TPR = \frac{TP}{TP + FN}$$

$$FPR = \frac{FP}{FP + TN}$$

predicted



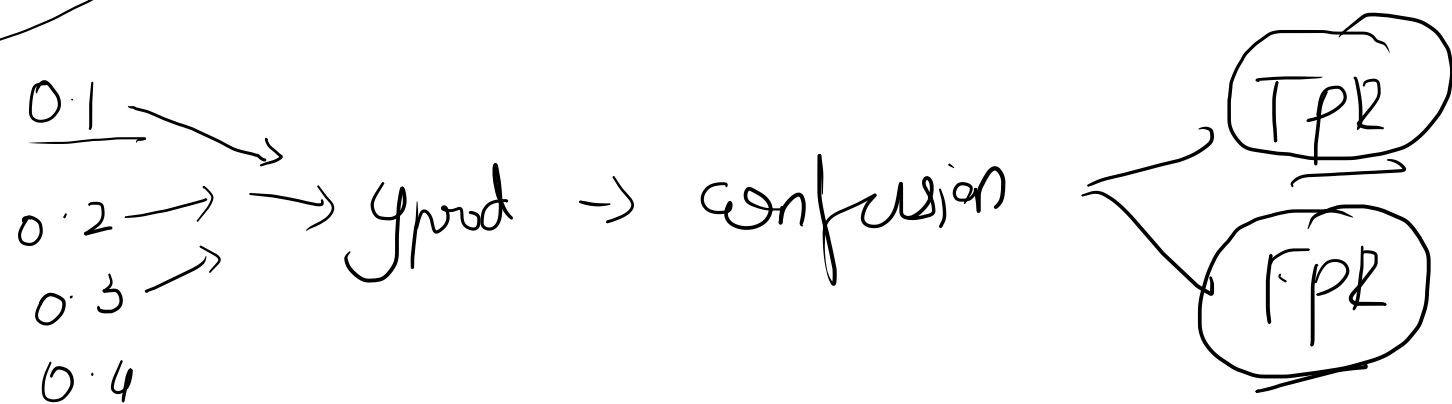
0.5

precision

recall

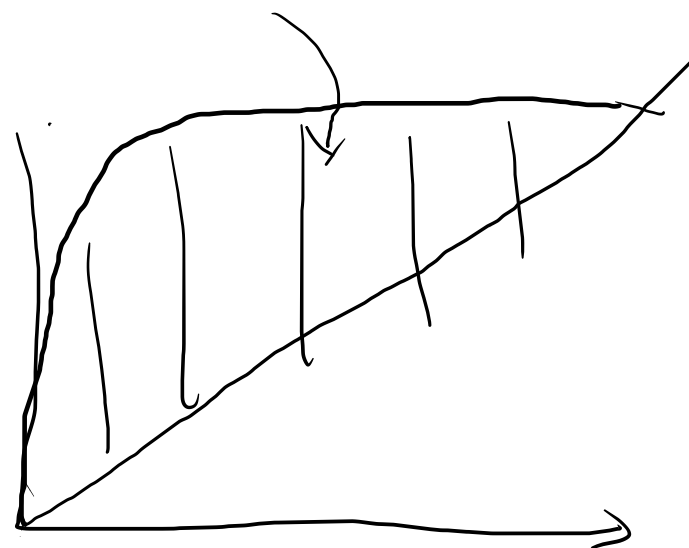
100

threshold \rightarrow y-prod \rightarrow confusion \rightarrow Accuracy



ROC

\uparrow TPR



FPR \downarrow

AUC \uparrow
 \rightarrow 0.5

Model

AUC 0.5

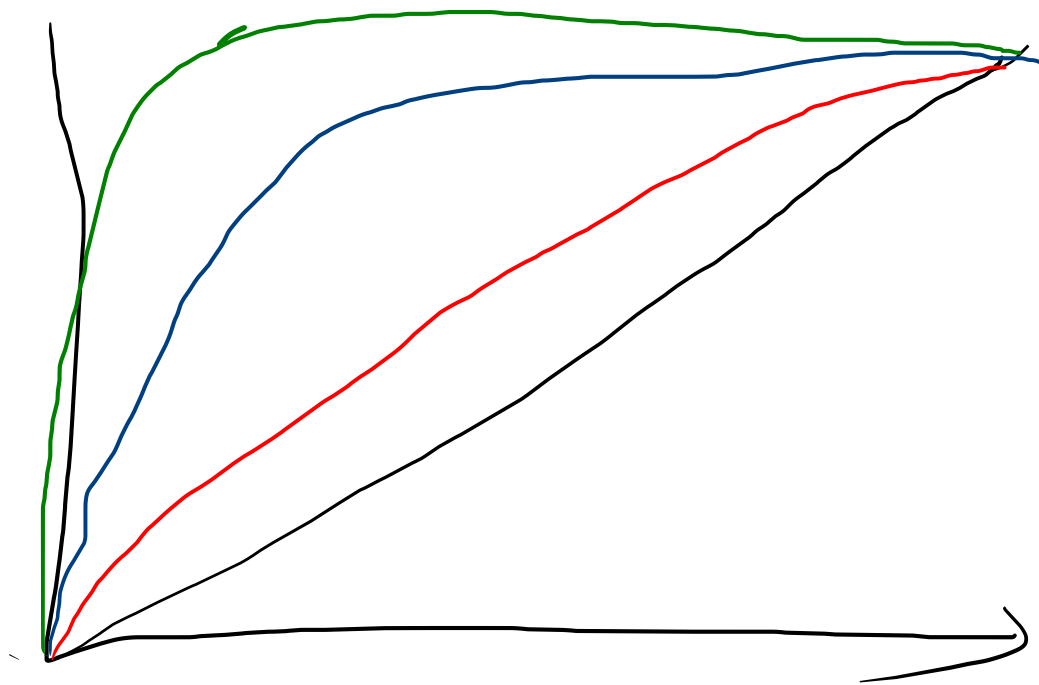
\rightarrow 0.60

\rightarrow 0.70

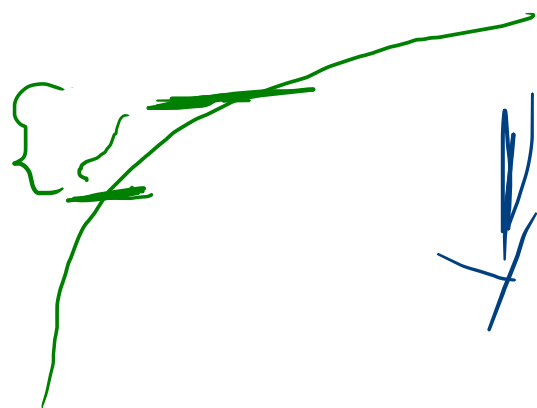
\rightarrow 0.80

best \Rightarrow 0.9-1

Model \rightarrow 0.5



green = ^{max} AUC



$$P \cdot R =$$

$$\frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}}$$