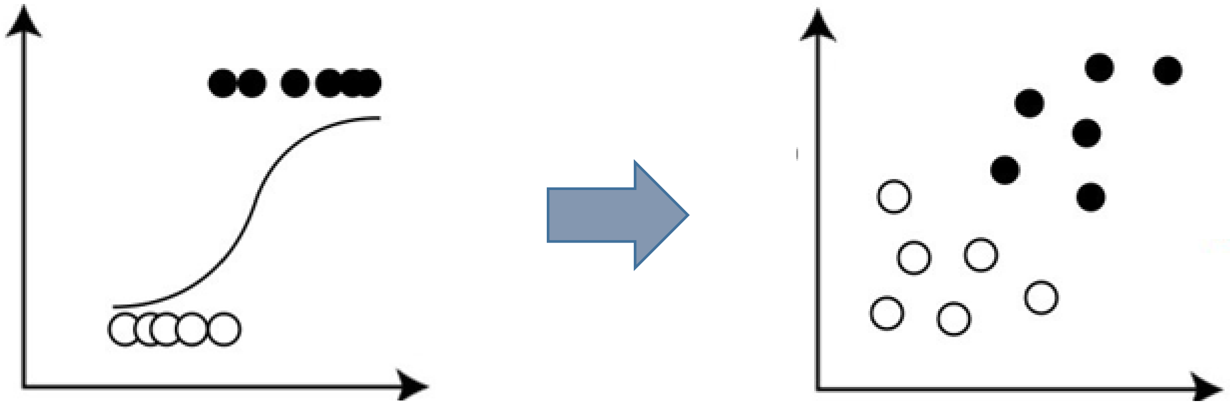
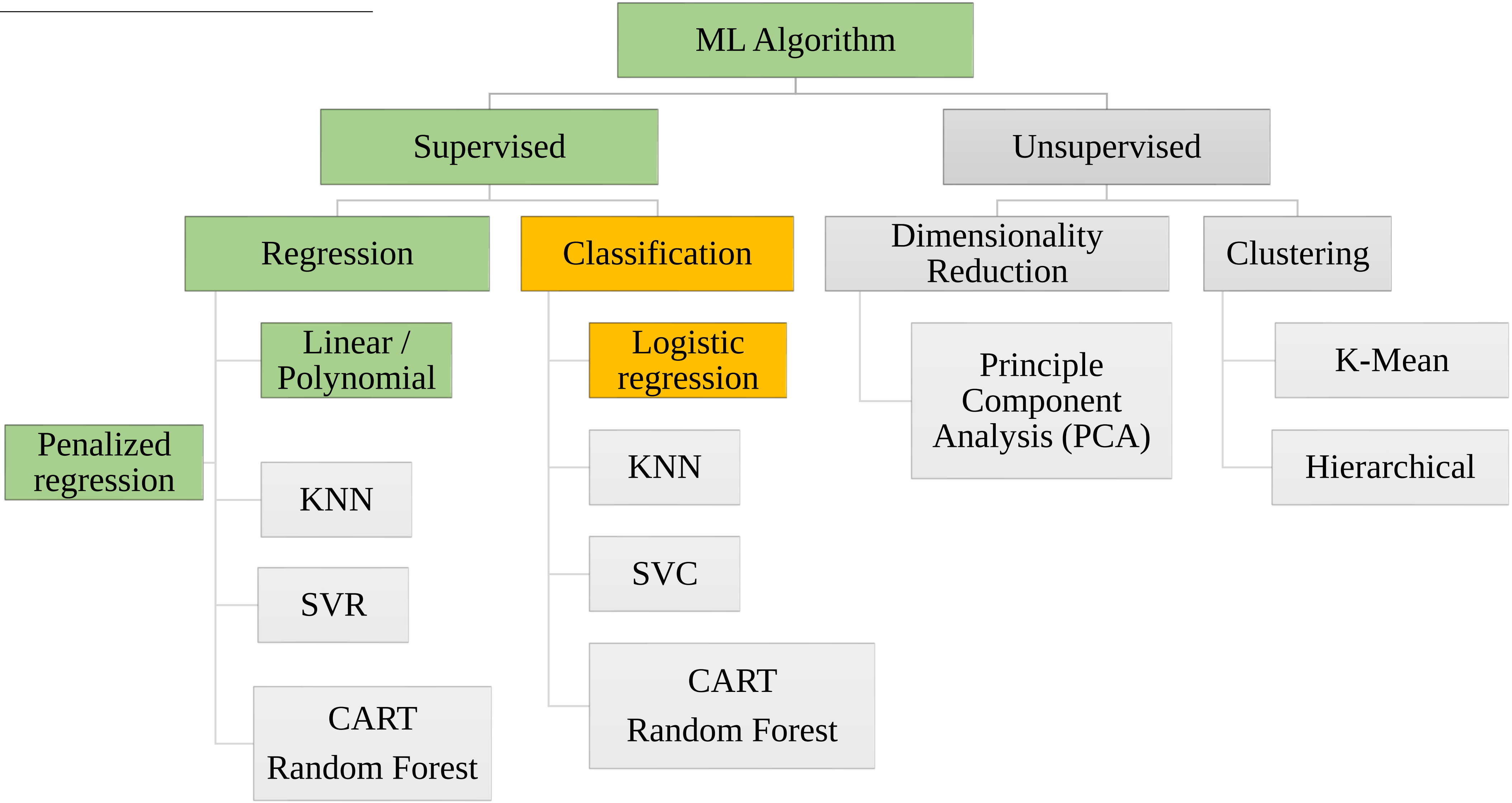
Class 10 – Logistic Regression

Prof. Pedram Jahangiry



Road map


ML Algorithm
Supervised
Unsupervised
Regression
Classification
Dimensionality Reduction
Clustering
Linear / Polynomial
Penalized regression
KNN
SVR
CART
Random Forest
Logistic regression
KNN
SVC
CART
Random Forest
Principle Component Analysis (PCA)
K-Mean
Hierarchical

Topics

<u>Part I</u>

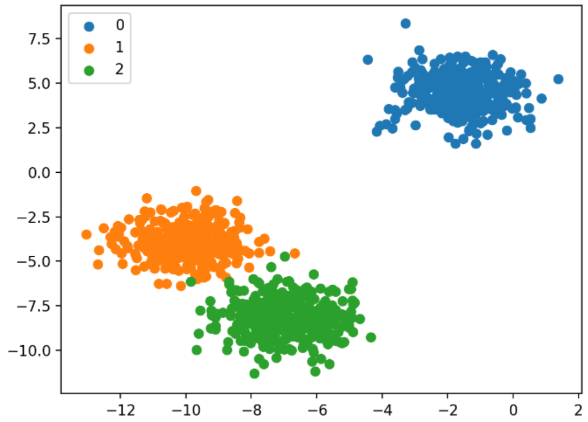
1. Linear probability model (LPM) vs Logistic regression
2. Sigmoid function
3. Logistic regression

<u>Part II</u>

1. Classification performance metrics
    a) Accuracy,
    b) Precision,
    c) Recall,
    d) F1 score,
    e) ROC and AUC.

		Predictions	
		0 negative	1 positive
Actual	0 negative	TN	FP
	1 positive	FN	TP

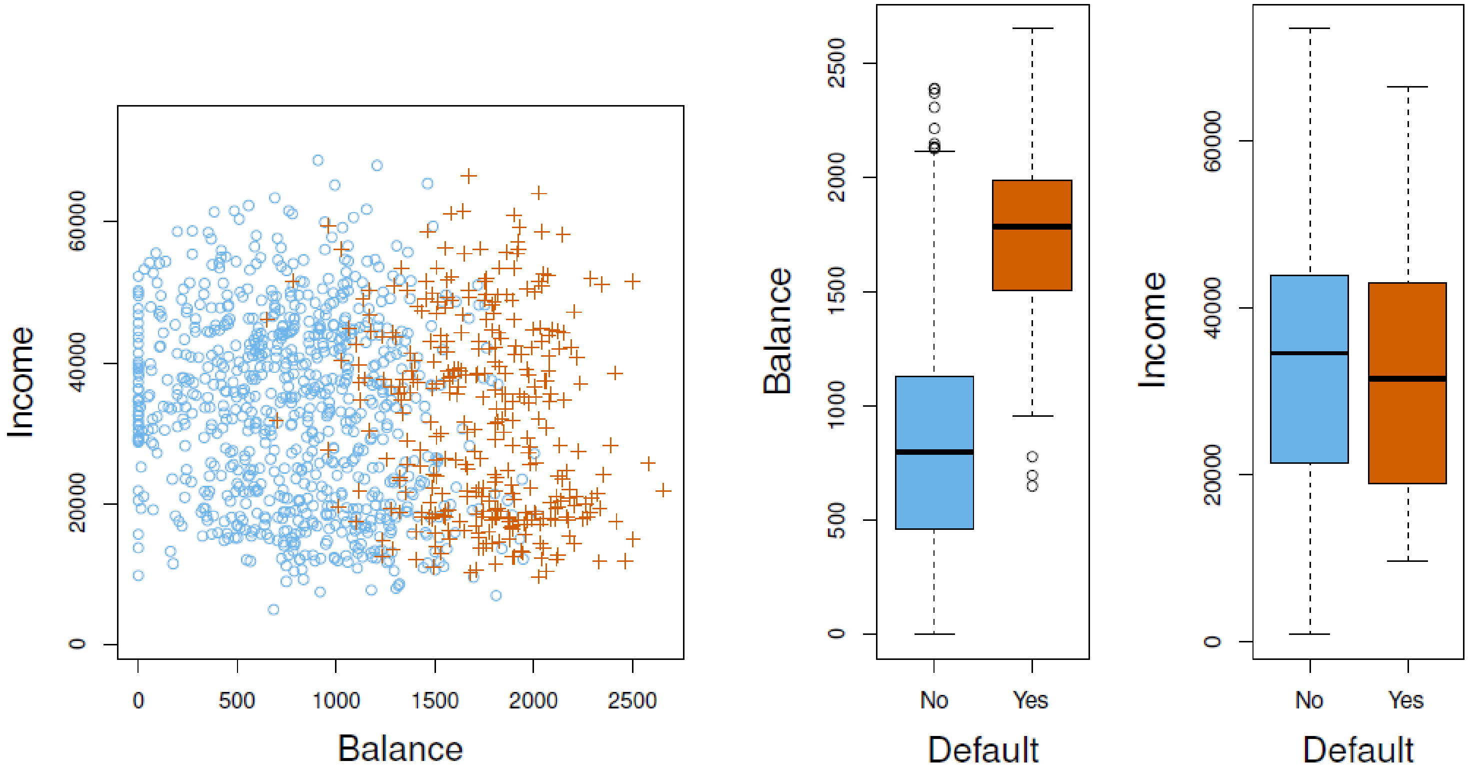
Classification



- Qualitative variables describe data that fits into categories.
- Qualitative variables are often referred to as **categorical.**
- **Classification** is the process of predicting categorical variables.
- Classification problems are quite common, perhaps even more than regression problems.
- **Examples:**
    - Online transactions (fraudulent or not)
    - Loan application (approval or denial)
    - Credit card default (default or not)
    - Car insurance customers (high, medium, low risk)

Credit card default example

- Goal: Build a classifier that performs well in both train and test set.



Part I

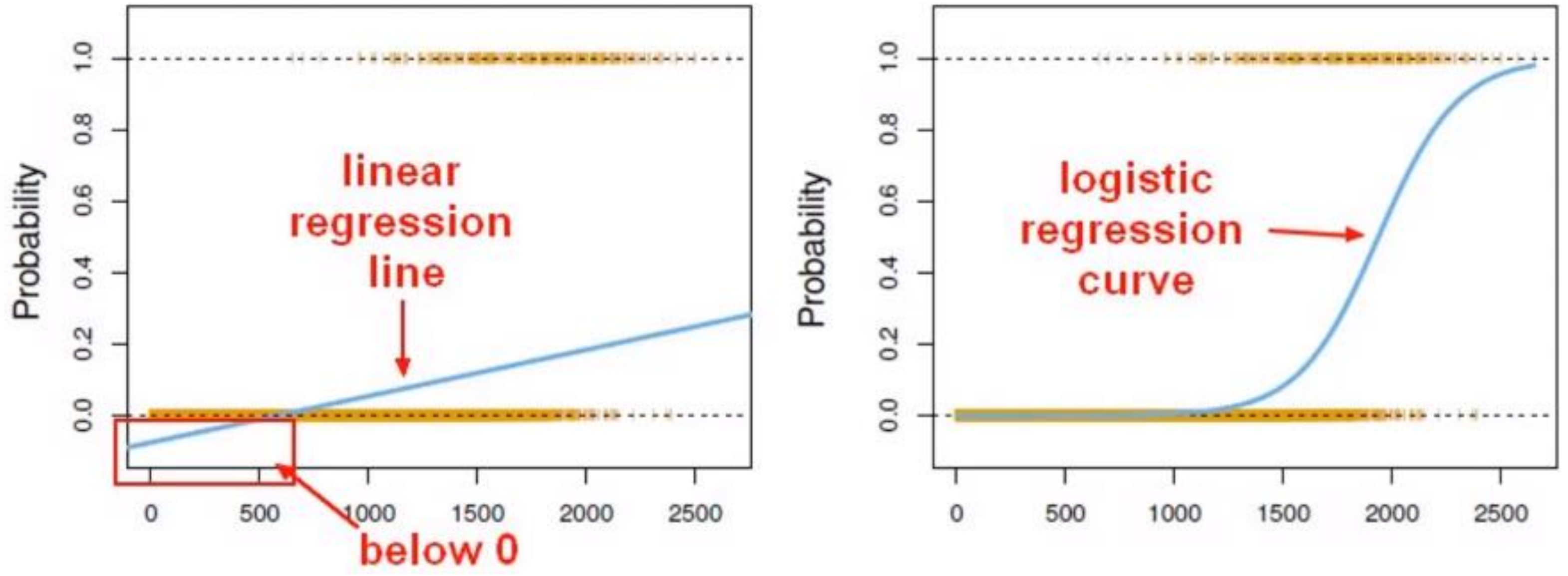
Logistic Regression

Linear Probability Model (LPM) vs Logistic Regression

Starting with simple LPM : $y = \beta_0 + \beta_1 bal + \epsilon$ where, $Y = 1$ for default and 0 otherwise.

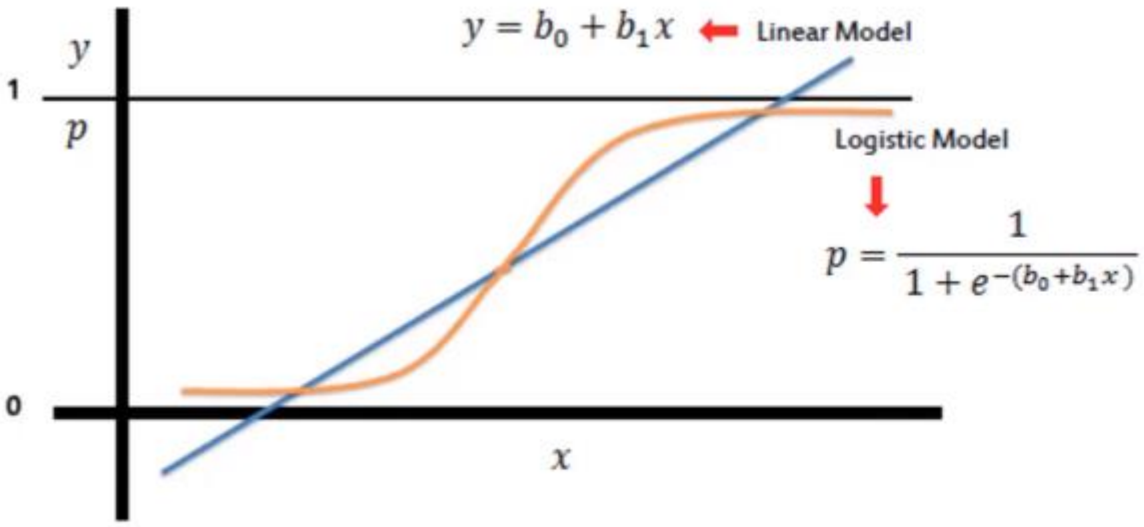
$$E(Y = 1|bal) = \Pr(Y = 1|bal) = P(x) = \beta_0 + \beta_1 bal$$

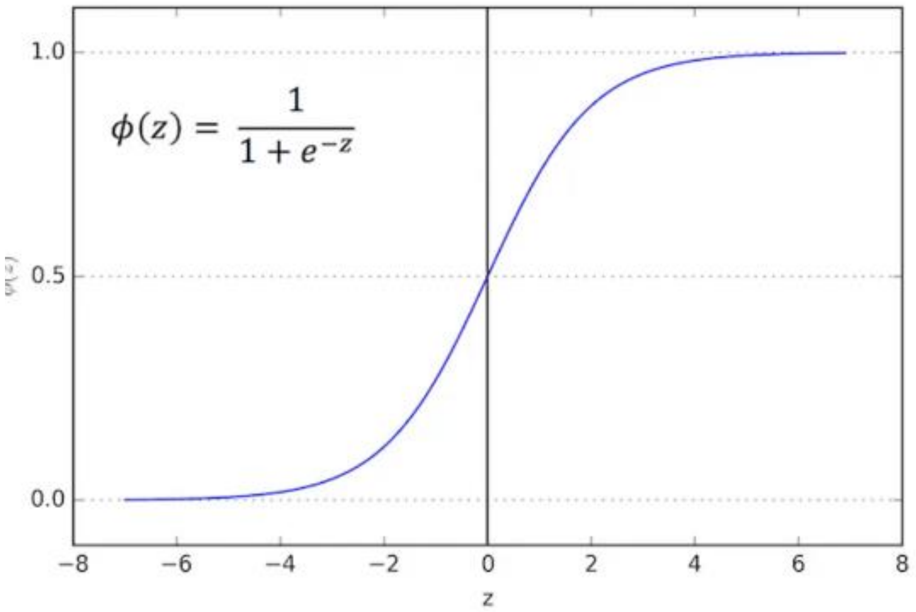
- It seems that simple regression is perfect for this task, but:



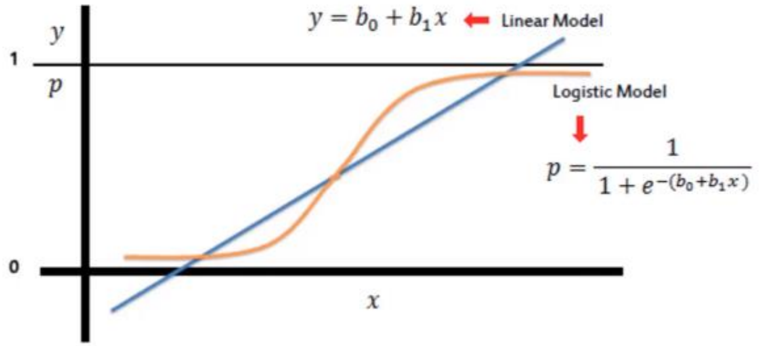
Sigmoid Function

- We need a monotone mapping function that has a range of [0,1]





Logistic Regression (Model)



- The model: $$f_{w,b}(X) = \frac{1}{1+e^{-(WX+b)}}$$

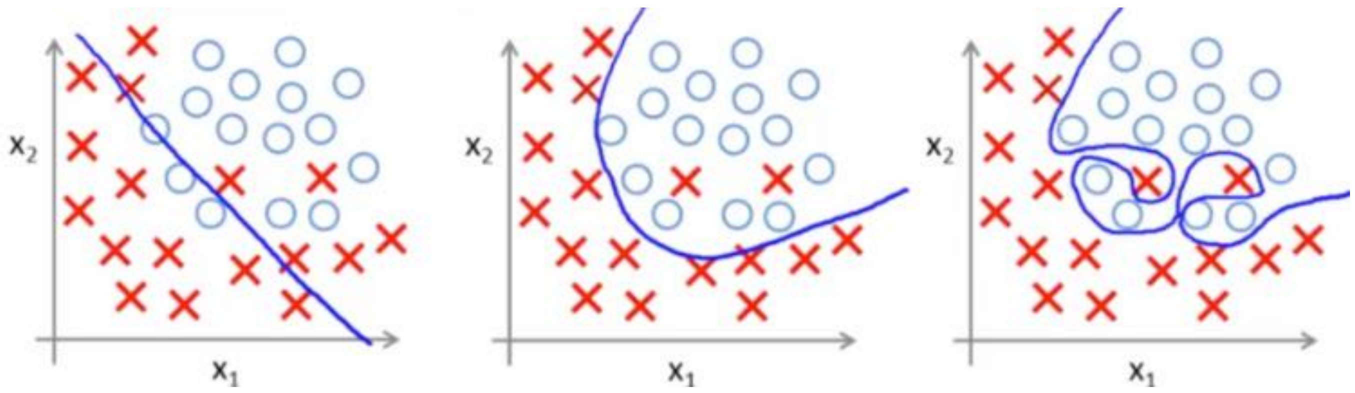
- In case of two classes, $f_{w,b}(X) = \Pr(Y = 1|x) = p(x)$.
- A bit of rearrangement gives

$$Log\left(\frac{p(X)}{1-p(x)}\right) = WX + b$$

- This monotone transformation is called the log odds or logit transformation of $p(x)$.
- Logistic regression ensures that our estimates always lie between 0 and 1

Logistic regression fit (Decision boundary)

- Depending on how we define $WX + b$, we can get any of the following fits from logistic regression classifier.



Logistic Regression (Objective function)

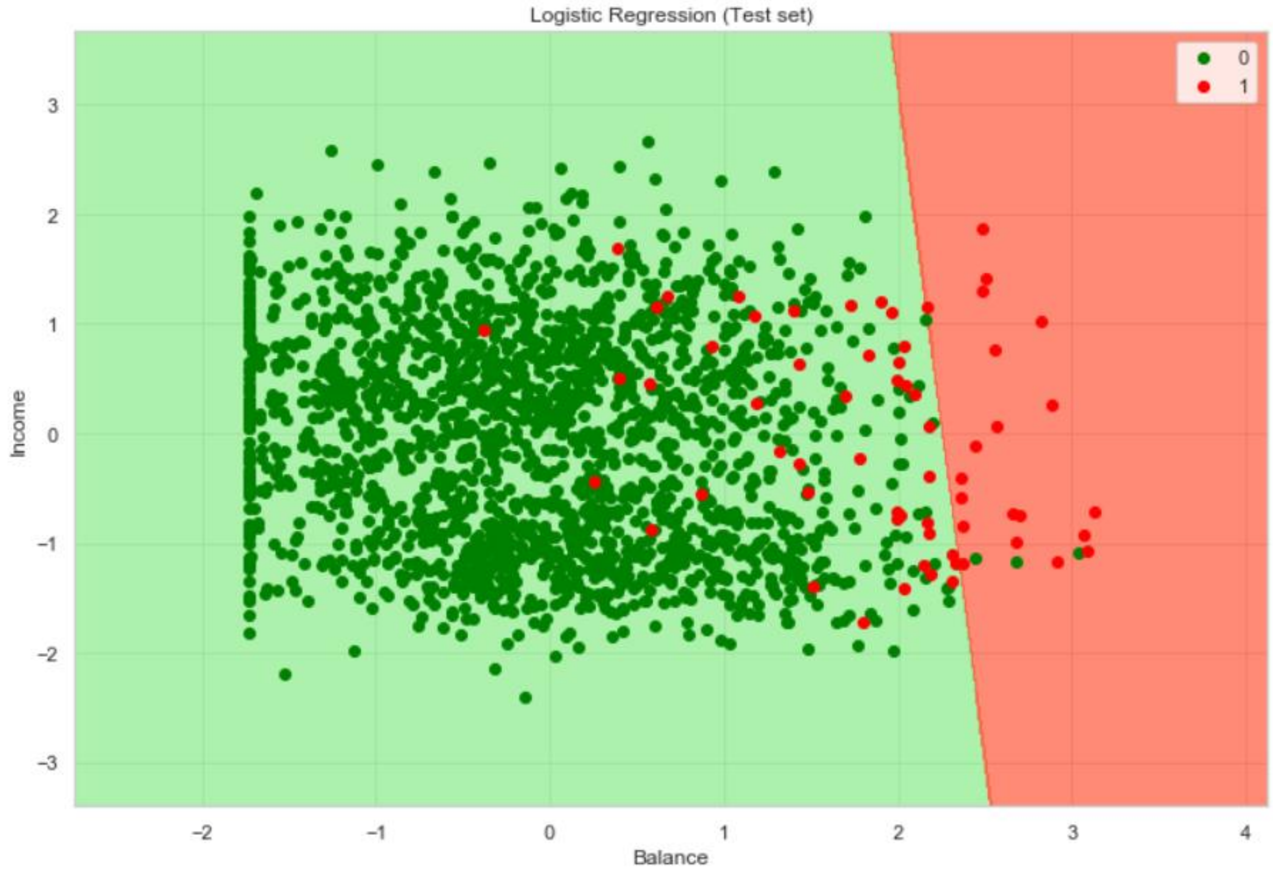
- In logistic regression, instead of minimizing the average loss, like in linear regression, we maximize the **likelihood** of the training data according to our model.

$$Max\ \{L_{w,b} = \prod_i \ f_{w,b}(x_i)^{y_i} \left(1 - f_{w,b}(x_i)\right)^{1-y_i} \}$$

- **Solution**: In practice, it is more convenient to maximize the log-likelihood function. This log-likelihood maximization, gives us w^* and b^*. We are now ready to make **predictions**.
- Depending on how we define the probability threshold, we can classify the observations. In practice, the <u>choice of the threshold</u> could be different depending on the problem.
- There is no closed form solution to this optimization problem. We need to use gradient descent.

Logistic regression output for credit card default example

$$P(default|bal, inc) = \frac{1}{1 + e^{\beta_0+\beta_1 bal+\beta_2 inc+\epsilon}}$$



		Predictions (Decision boundary) 0 No Default	1 Default
Actual	0 No Default	TN=1933	FP=3
	1 Default	FN=44	TP=20

Part II
Classification Performance Metrics

Confusion Matrix

		Predictions	
		0 negative	1 positive
Actual	0 negative	TN	FP*
	1 positive	FN**	TP

FP* Type I error
FN** Type II eror

		predicted class		
		class 1	class 2	class 3
actual class	class 1	True positives		
	class 2		True positives	
	class 3			True positives

Accuracy, Precision, Recall and F1score

		Predictions	
		0 negative	1 positive
Actual	0 negative	TN	FP
	1 positive	FN	TP

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

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

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

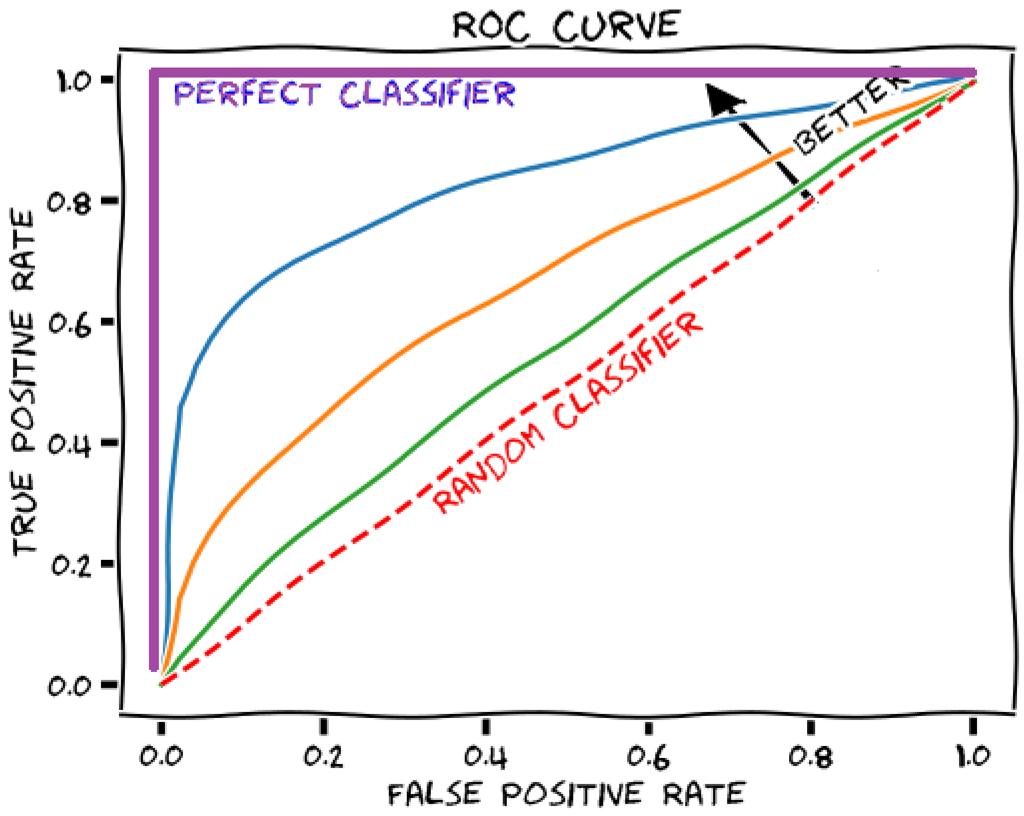
$$F1\ Score = 2 * \frac{PR}{P + R}$$

ROC (Receiver Operating Characteristic)

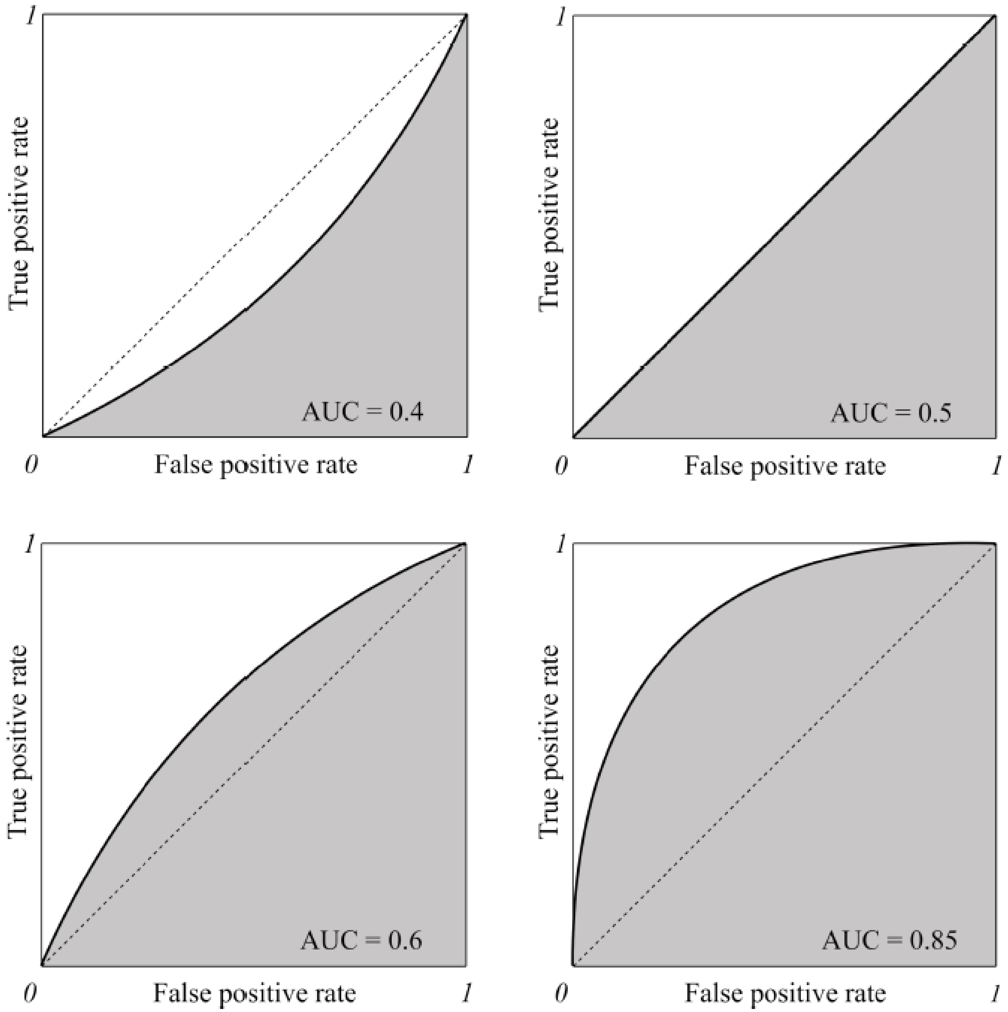
		Predictions	
		0 negative	1 positive
Actual	0 negative	TN	FP
	1 positive	FN	TP

$$False\ Positive\ Rate = \frac{FP}{FP + TN}$$

$$True\ Positive\ Rate = \frac{TP}{TP + FN}$$



AUC


1
True positive rate
AUC = 0.4
0
False positive rate
1
1
True positive rate
AUC = 0.5
0
False positive rate
1
1
True positive rate
AUC = 0.6
0
False positive rate
1
1
True positive rate
AUC = 0.85
0
False positive rate
1

Some other classification metrics

		True condition				
	Total population	Condition positive	Condition negative	Prevalence $= \frac{\Sigma \text{ Condition positive}}{\Sigma \text{ Total population}}$	Accuracy (ACC) = $\frac{\Sigma \text{ True positive} + \Sigma \text{ True negative}}{\Sigma \text{ Total population}}$	
Predicted condition	Predicted condition positive	**True positive**	**False positive**, Type I error	Positive predictive value (PPV), Precision = $\frac{\Sigma \text{ True positive}}{\Sigma \text{ Predicted condition positive}}$	False discovery rate (FDR) = $\frac{\Sigma \text{ False positive}}{\Sigma \text{ Predicted condition positive}}$	
	Predicted condition negative	**False negative**, Type II error	**True negative**	False omission rate (FOR) = $\frac{\Sigma \text{ False negative}}{\Sigma \text{ Predicted condition negative}}$	Negative predictive value (NPV) = $\frac{\Sigma \text{ True negative}}{\Sigma \text{ Predicted condition negative}}$	
		True positive rate (TPR), Recall, Sensitivity, probability of detection, Power $= \frac{\Sigma \text{ True positive}}{\Sigma \text{ Condition positive}}$	False positive rate (FPR), Fall-out, probability of false alarm $= \frac{\Sigma \text{ False positive}}{\Sigma \text{ Condition negative}}$	Positive likelihood ratio (LR+) $= \frac{\text{TPR}}{\text{FPR}}$	Diagnostic odds ratio (DOR) $= \frac{\text{LR+}}{\text{LR−}}$	F_1 score = $2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$
		False negative rate (FNR), Miss rate $= \frac{\Sigma \text{ False negative}}{\Sigma \text{ Condition positive}}$	Specificity (SPC), Selectivity, True negative rate (TNR) $= \frac{\Sigma \text{ True negative}}{\Sigma \text{ Condition negative}}$	Negative likelihood ratio (LR−) $= \frac{\text{FNR}}{\text{TNR}}$		

Students' questions