Classification

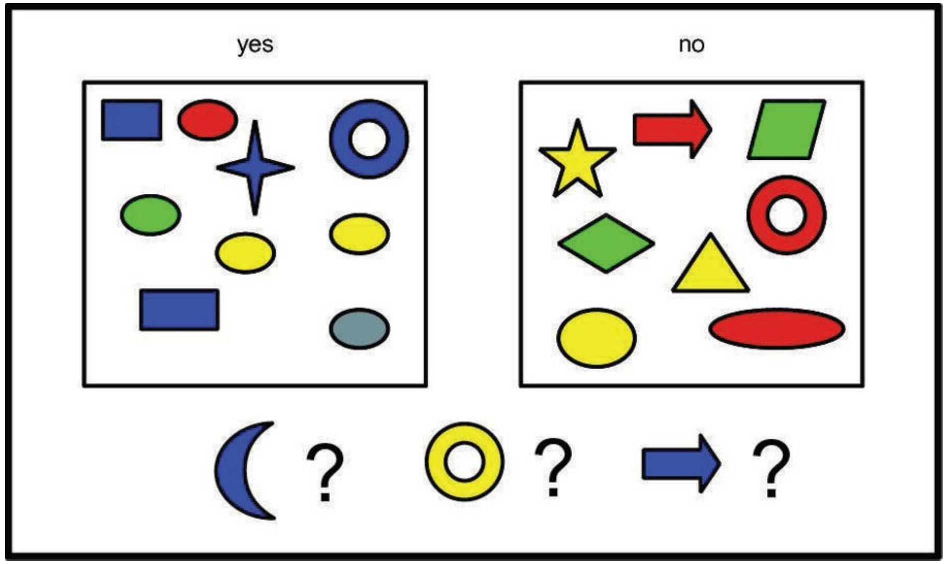
JinYeong Bak

jy.bak@skku.edu

Human Language Intelligence Lab, SKKU

Supervised Learning

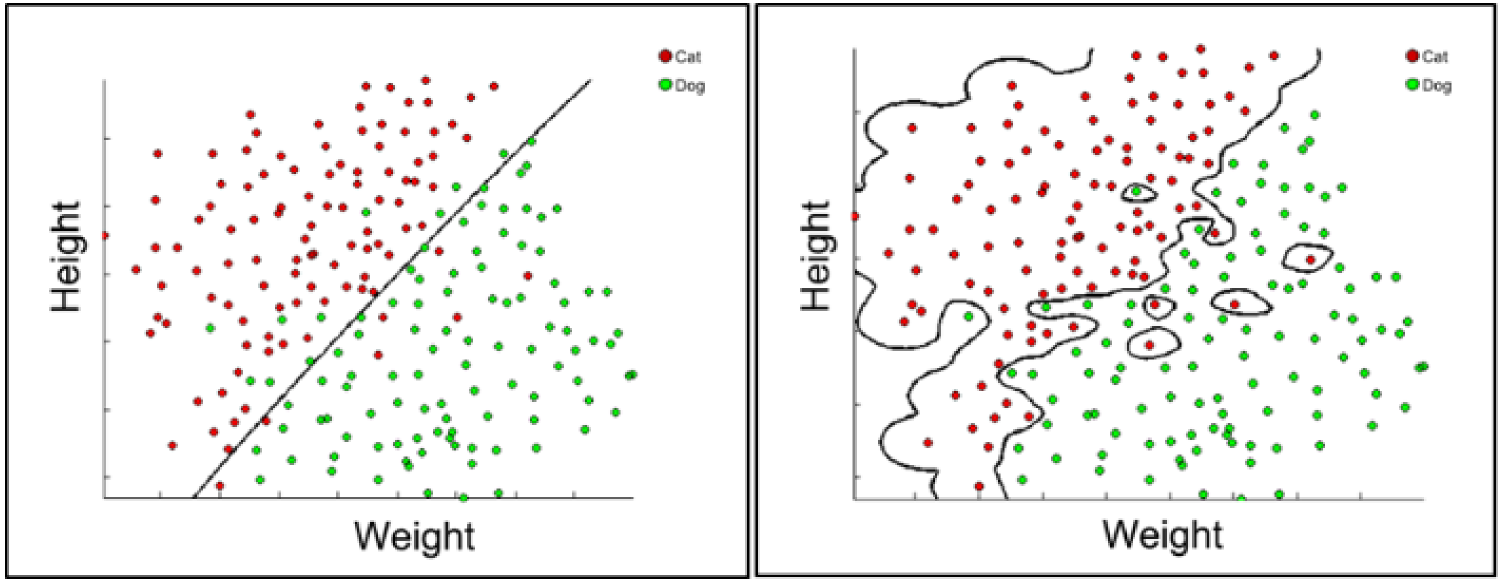
- Given: Training data as labeled instances $\{(x^{(1)}, y^{(1)}), \dots, (x^{(N)}, y^{(N)})\}$
- Goal: Learn a rule $(f: x \rightarrow y)$ to predict outputs y for new inputs x
- Example)
  - Data: ((Blue, Square, 10), yes), . . . ((Red, Ellipse, 20.7), yes)
  - Task: For new inputs (Blue, Crescent, 10), (Yellow, Circle, 12), are they yes/no?



Color	Shape	Size	Label
Blue	Square	10	1
Red	Ellipse	2.4	1
Red	Ellipse	20.7	0
Blue	Crescent	10	?
Yellow	Circle	12	?

Supervised Learning

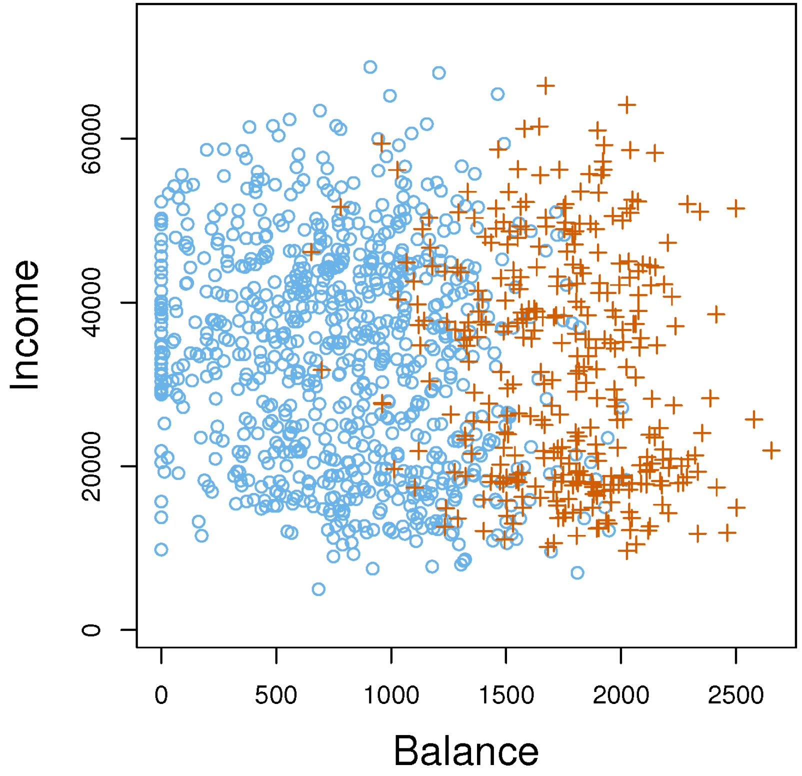
- Classification: Discrete-valued outputs
- Example)
  - Data: Size and label {(Height, Weight), Cat/Dog}
  - Task: Predict whether an animal is a cat or dog given new size information
  - Method: Finding a linear or nonlinear separator

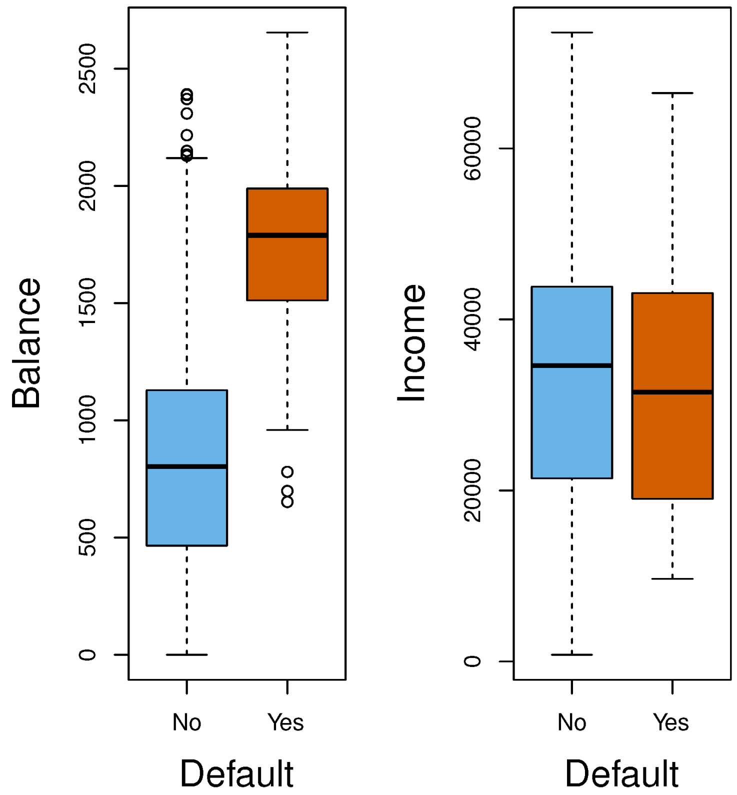


LOGISTIC REGRESSION

Problem

- Data: Credit card balance, annual income, default or not
  {(Balance, Income), Default?}
- Task: Predict whether a person will default on his/her credit card payment





Data

- N: # training data
- X_1, X_2: Balance, Income
- Y: Default or not
- (x, y): one training data
- $(x_1^{(i)}, x_2^{(i)}, y^{(i)})$: i-th training data

X_1	X_2	Y
729.52	44361.62	No
817.18	12106.13	No
1570.65	16239.15	Yes
529.25	35704.49	No
785.65	38463.49	No
1321.53	23735.15	Yes
1377.68	41435.26	Yes
⋮	⋮	⋮

Linear Regression

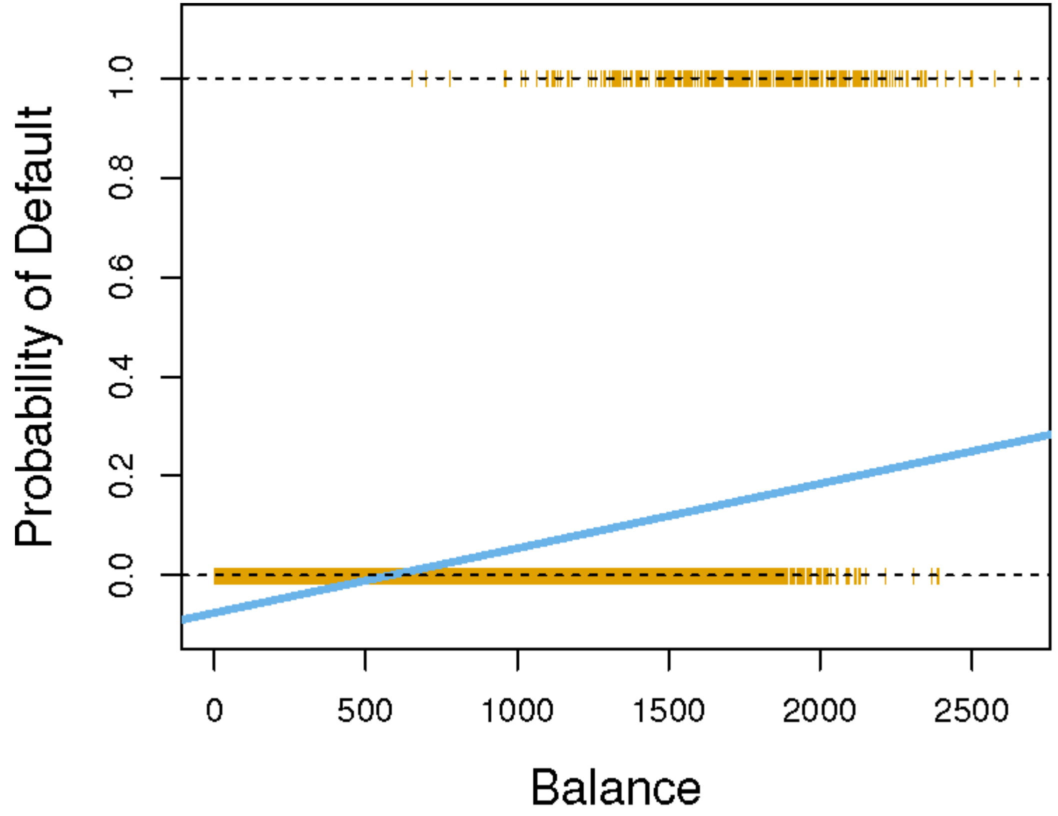
- Data: {(Balance, Income), Default?}
- Task: Predict default $y^{(test)}$ based on income and balance $x_1^{(test)}, x_2^{(test)}$
- Model: $Y \approx \beta_0 + \beta_1 X_1 + \beta_2 X_2$
- Problem: Y is not a number, just yes/no
  - The number has properties such as order and gap between elements
    ex) 1 < 2, 3 – 1 = 6 – 4
  - No ordering nor gap between qualitative response
    ex) yes < no?, yes – no??

Linear Regression

- Data: {(Balance, Income), Default?}
- Task: Predict default $y^{(test)}$ based on income and balance $x_1^{(test)}, x_2^{(test)}$
- Model: $Y \approx \beta_0 + \beta_1 X_1 + \beta_2 X_2$
- Problem: Y is not a number, just yes/no
- Solution: Use the probability
  - We want to know whether Y is yes or no
  - $P(Y = yes)$
  - Cases
    - $P(Y = yes) \approx 1$: We can say that Y is yes
    - $P(Y = yes) \approx 0$: We can say that Y is no
    - $P(Y = yes) \approx 0.4$: Hmm... we might say that Y is no since the probability is less than half

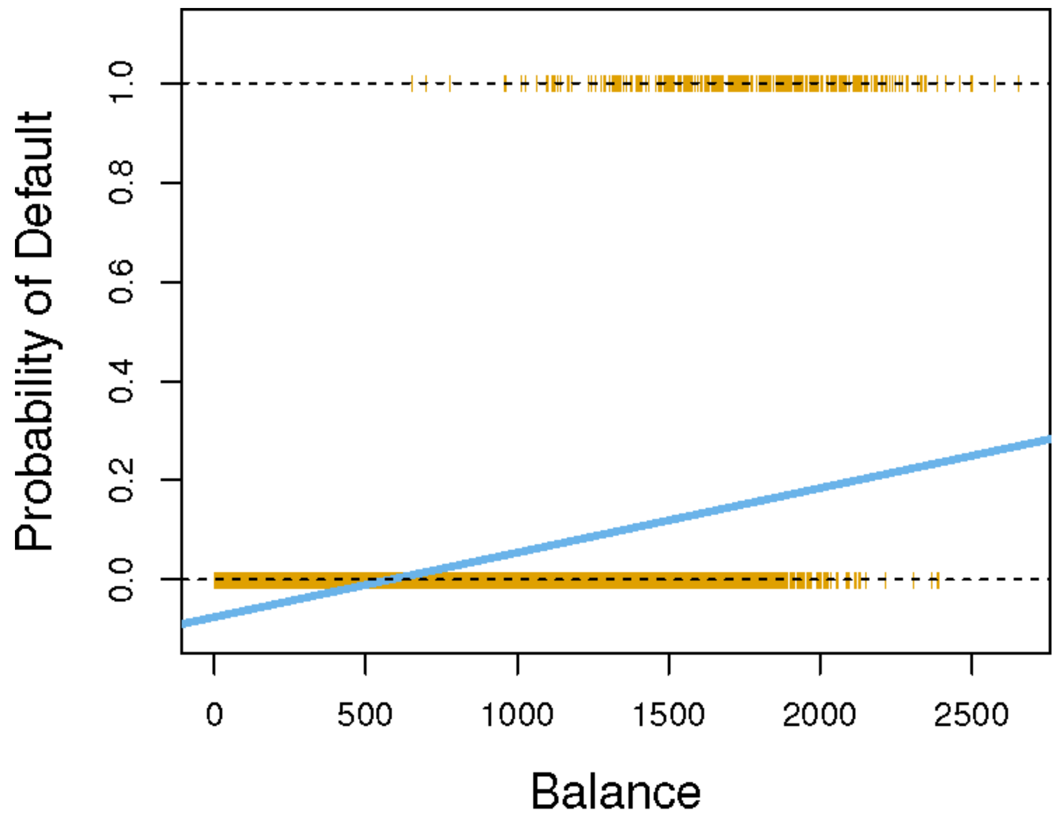
Linear Regression

- Data: {(Balance, Income), Default?}
- Task: Predict default $y^{(test)}$ based on income and balance $x_1^{(test)}, x_2^{(test)}$
- Model: $Y \approx \beta_0 + \beta_1 X_1 + \beta_2 X_2$
- Problems
  - Existence of $P(Y = yes) < 0$
  - No existence of $P(Y = yes) > 0.5$



Linear Regression

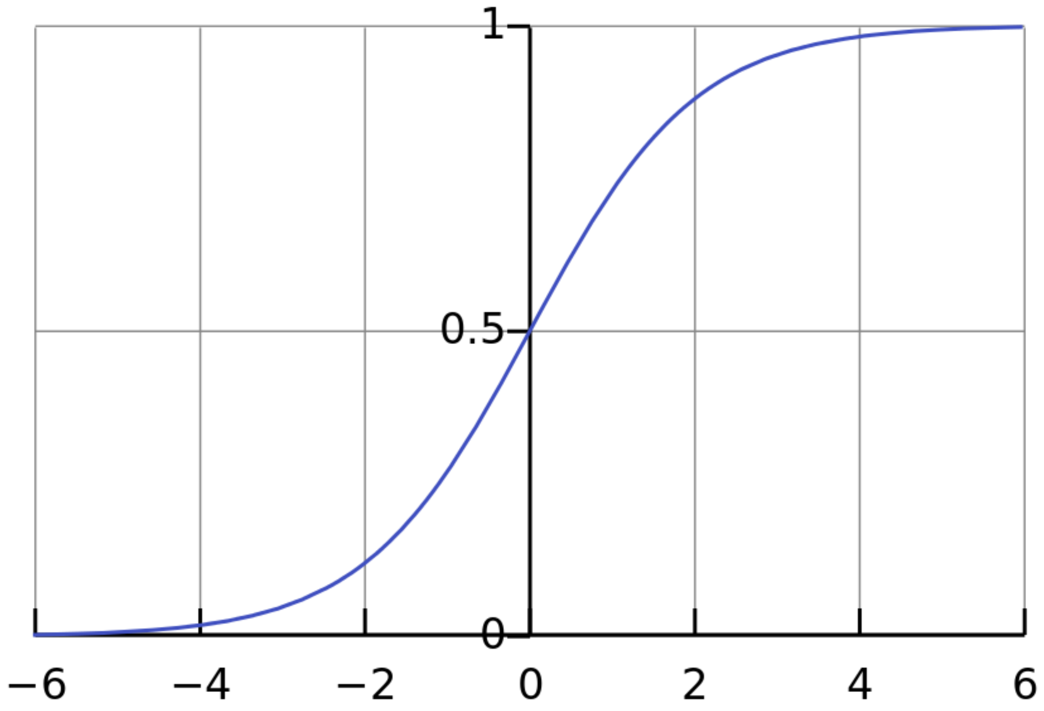
- Data: {(Balance, Income), Default?}
- Task: Predict default $y^{(test)}$ based on income and balance $x_1^{(test)}, x_2^{(test)}$
- Model: $Y \approx \beta_0 + \beta_1 X_1 + \beta_2 X_2$
- Problems
  - Existence of $P(Y = yes) < 0$
  - No existence of $P(Y = yes) > 0.5$
- Solution: Limit the range of Y
  - Use a function to change the range
  - $-\infty < Y < \infty$ => $0 \leq P(Y = yes) \leq 1$



Logistic Function

$$g(x) = \frac{1}{1+e^{-x}} = \frac{e^x}{1+e^x}$$

$$\frac{d}{dx}g(x) = g(x)\big(1-g(x)\big), 1-g(x) = g(-x)$$



Logistic Regression

- Data: {(Balance, Income), Default?}
- Task: Predict default $y^{(test)}$ based on income and balance $x_1^{(test)}, x_2^{(test)}$
- Model: $\mathrm{P}(Y = yes) \approx g(\beta_0 + \beta_1 X_1 + \beta_2 X_2)$

Logistic Regression

- Data: {(Balance, Income), Default?}
- Task: Predict default $y^{(test)}$ based on income and balance $x_1^{(test)}, x_2^{(test)}$
- Model: $\mathrm{P}(Y = yes) \approx g(\beta_0 + \beta_1 X_1 + \beta_2 X_2)$
- Decision boundary
  - $P(Y = yes) \geq 0.5$: yes
  - $P(Y = yes) < 0.5$: no

Logistic Regression

- Data: {(Balance, Income), Default?}
- Task: Predict default $y^{(test)}$ based on income and balance $x_1^{(test)}, x_2^{(test)}$
- Model: $\mathrm{P}(Y = yes) \approx g(\beta_0 + \beta_1 X_1 + \beta_2 X_2)$
- Idea: Given a data $x^{(i)}$, minimize the difference between $\hat{y}^{(i)}$ and $y^{(i)}$
  - $\hat{y}^{(i)}$: output of the model with β_0, β_1 and β_2
  - $y^{(i)}$: real data output (yes: 1, no: 0)
- Difference: $\left(y^{(i)} - \hat{y}^{(i)}\right)^2 = \left(y^{(i)} - g\left(\beta_0 + \beta_1 x_1^{(i)} + \beta_2 x_2^{(i)}\right)\right)^2$
- Method: Find the best β_0, β_1 and β_2 that minimize the all data difference

$$\underset{\beta_0,\beta_1,\beta_2}{\arg\min} \sum_{i}^{N} \left(y^{(i)} - g\left(\beta_0 + \beta_1 x_1^{(i)} + \beta_2 x_2^{(i)}\right)\right)^2$$

Cannot use gradient-descent algorithm since it is non-convex

Maximum Likelihood Estimation

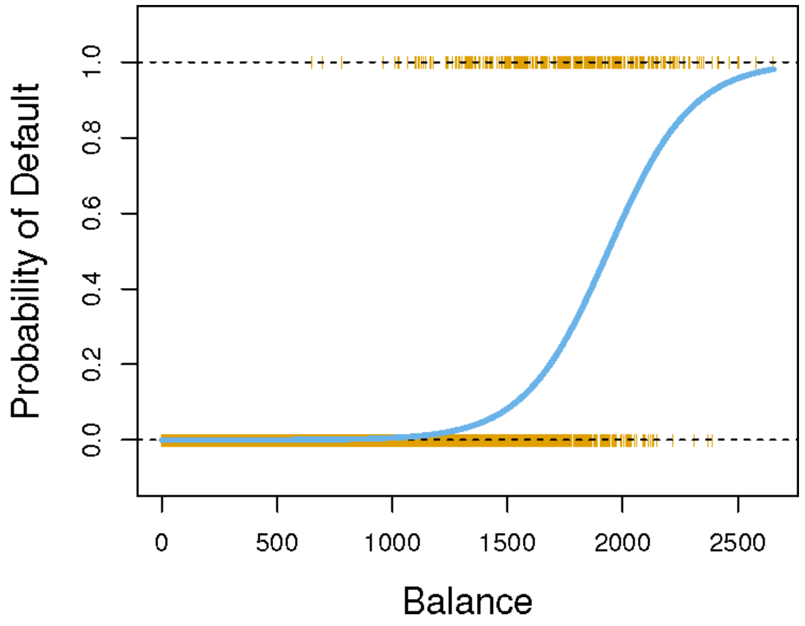
- Likelihood function
  - $P(Y^{(i)}|X^{(i)}, \beta)$
  - Probability of outcome ($Y^{(i)}$) when the model parameters are (β) and input is ($X^{(i)}$)
- Maximum Likelihood Estimation
  - $\arg max_{\beta} P(Y^{(i)}|X^{(i)}, \beta)$
  - Find the parameters that maximize the probability of outcome data from the model that has the parameters and input data

Maximum Likelihood Estimation

- Logistic regression
  - Maximize the likelihood function
  - Find the parameters that maximize the probability of outcome data from the model that has the parameters and input data
- Linear regression
  - Minimize the loss function
  - Find the parameters that minimize the difference between model output and outcome data
- Negative likelihood function $\approx$ loss function

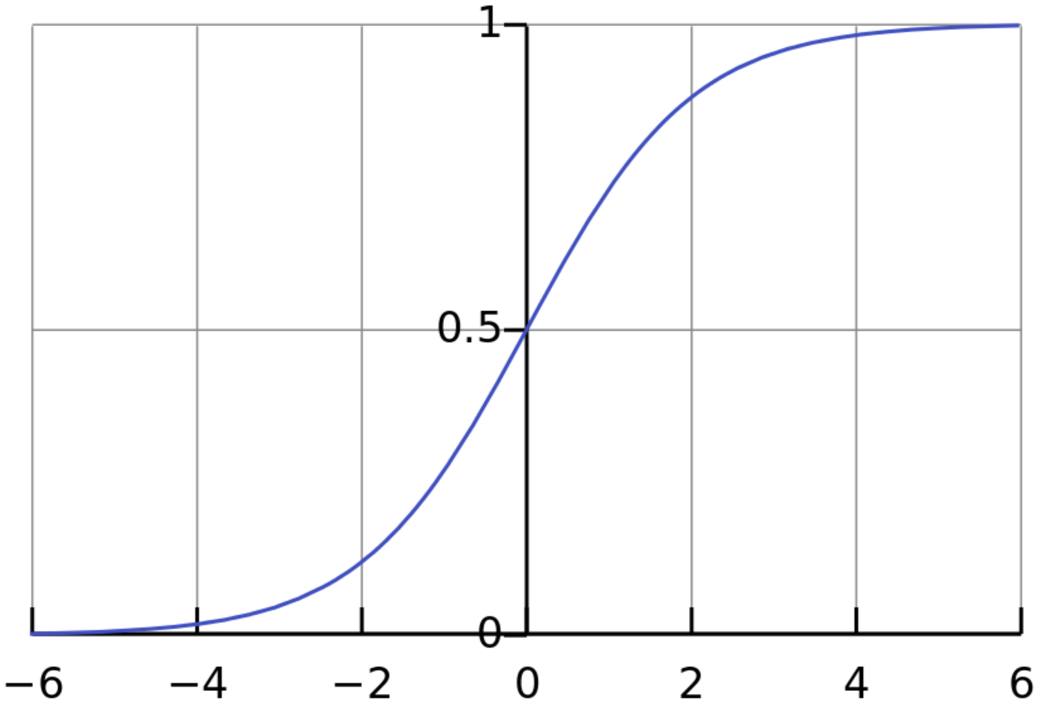
Logistic Regression

- Data: {(Balance, Income), Default?}
- Task: Predict default $y^{(test)}$ based on income and balance $x_1^{(test)}, x_2^{(test)}$
- Model: $\mathrm{P}(Y = yes) \approx g(\beta_0 + \beta_1 X_1 + \beta_2 X_2)$
- Method: Find the best β_0, β_1 and β_2 that maximize the likelihood function
- Algorithm: gradient-descent algorithm



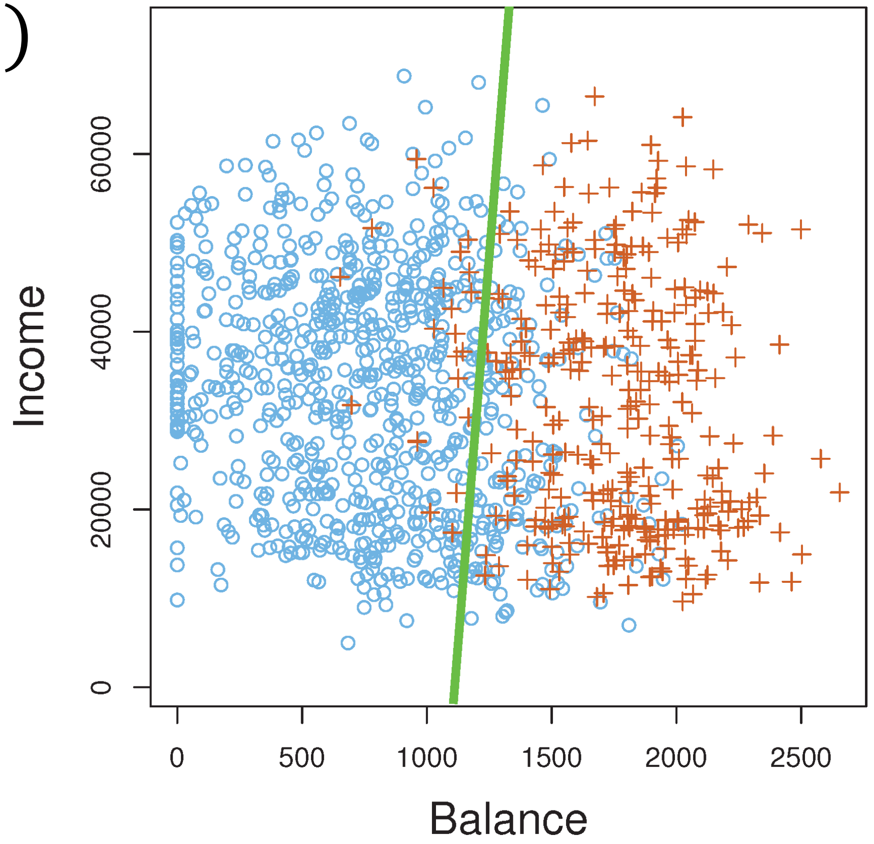
Logistic Regression

- Data: {(Balance, Income), Default?}
- Task: Predict default $y^{(test)}$ based on income and balance $x_1^{(test)}, x_2^{(test)}$
- Model: $\mathrm{P}(Y = yes) \approx g(\beta_0 + \beta_1 X_1 + \beta_2 X_2)$
- Decision boundary
  - $P(Y = yes) \geq 0.5$: yes
  - $P(Y = yes) < 0.5$: no
- Another decision boundary (log-odds)
  - $\beta_0 + \beta_1 X_1 + \beta_2 X_2 \geq 0$: yes
  - $\beta_0 + \beta_1 X_1 + \beta_2 X_2 < 0$: no



Logistic Regression

- Data: {(Balance, Income), Default?}
- Task: Predict default $y^{(test)}$ based on income and balance $x_1^{(test)}, x_2^{(test)}$
- Model: $\mathrm{P}(Y = yes) \approx g(\beta_0 + \beta_1 X_1 + \beta_2 X_2)$
- Decision boundary: $\beta_0 + \beta_1 X_1 + \beta_2 X_2$



Supervised Learning

- Problem: Predict outputs y for new inputs x based on a rule $(f: x \rightarrow y)$
- Data: Labeled instances $\{(x^{(1)}, y^{(1)}), \dots, (x^{(N)}, y^{(N)})\}$
- Model: Supervised model (e.g. linear regression, logistic regression)
- Parameters: Unknown values of the model
- Loss function: Difference between the outputs of the model and the data
- Task: Find the parameters that minimize the loss function
- Algorithm: Various algorithms

CLASSIFICATION PERFORMANCE

Classification Performance

- Questions
  - Which model will be best?
  - How to measure the performance of classification model?
- Answer: Confusion matrix

		Predicted	
		Yes	No
Actual	Yes	True Positive (TP)	False Negative (FN)
	No	False Positive (FP)	True Negative (TN)

Classification Performance - Measurement

- Accuracy
  - Did the model get it right?
  - $(TP + TN)/ALL$
- Precision
  - How many selected items are relevant?
  - $TP/Prected\ "yes"$
- Recall
  - How many relevant items are selected?
  - $TP/Actual\ "yes"$
- F score
  - Combination of precision and recall
  - $2 * (Precision\ * Recall)/(Precision + Recall)$

		Predicted	
		Yes	No
Actual	Yes	True Positive (TP)	False Negative (FN)
	No	False Positive (FP)	True Negative (TN)

Classification Performance - Measurement

- Example) Predict whether a person will default, Logistic regression, 100 test data
- Accuracy
  - Did the model get it right?
  - $(TP + TN)/ALL$
- Precision
  - How many selected items are relevant?
  - $TP/Prected\ "yes"$
- Recall
  - How many relevant items are selected?
  - $TP/Actual\ "yes"$
- F score
  - Combination of precision and recall
  - $2 * (Precision\ * Recall)/(Precision + Recall)$

		Predicted	
		Yes	No
Actual	Yes	70	15
	No	10	5