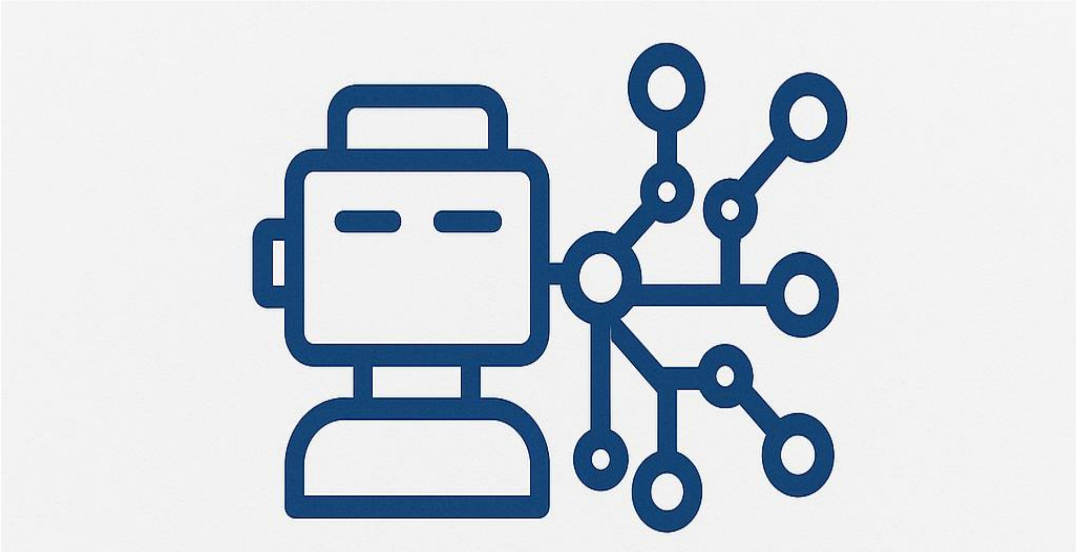
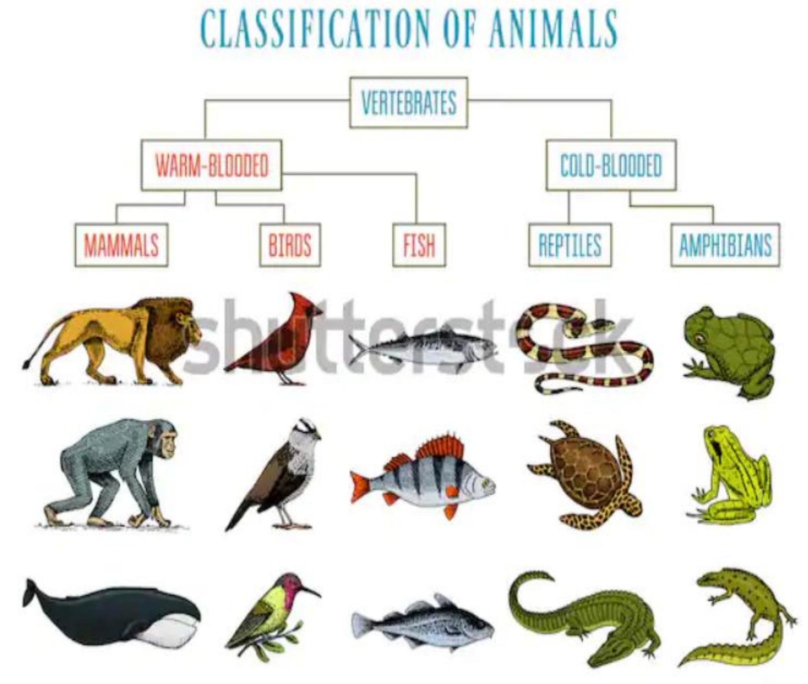


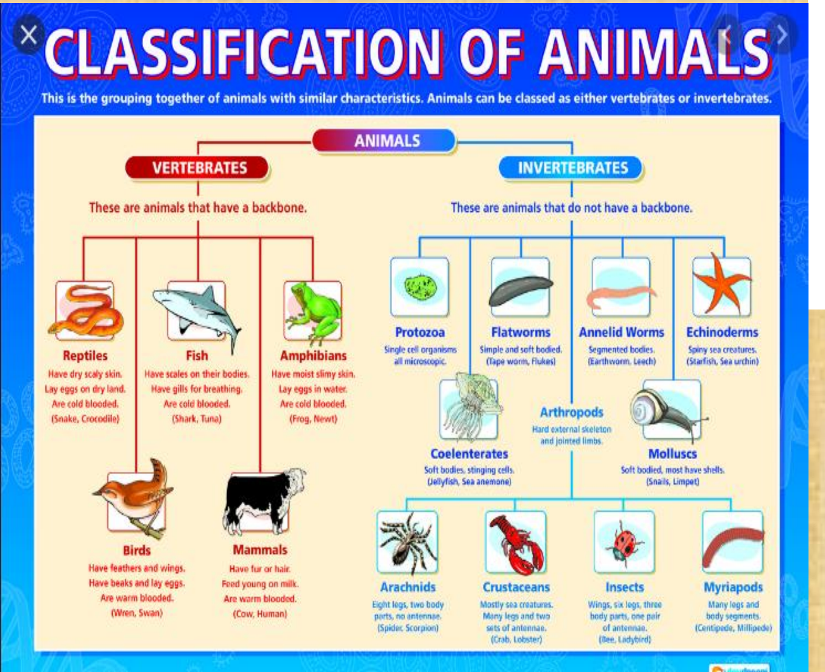
LOGISTIC REGRESSION PART 1

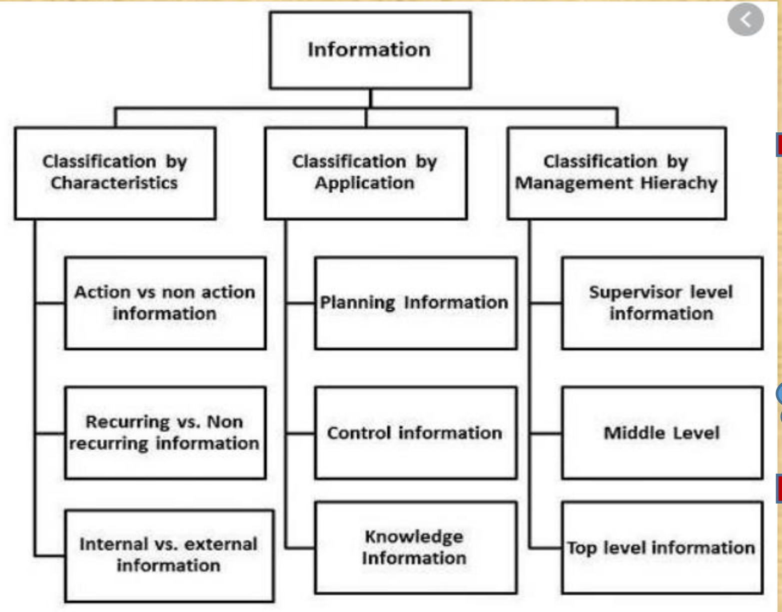
Ahmed Adil Nafea

Email: ahmed.a.n@uoanbar.edu.iq
Department of Artificial Intelligence, College of Computer Science and IT, University of Anbar, Ramadi 31001, Iraq.

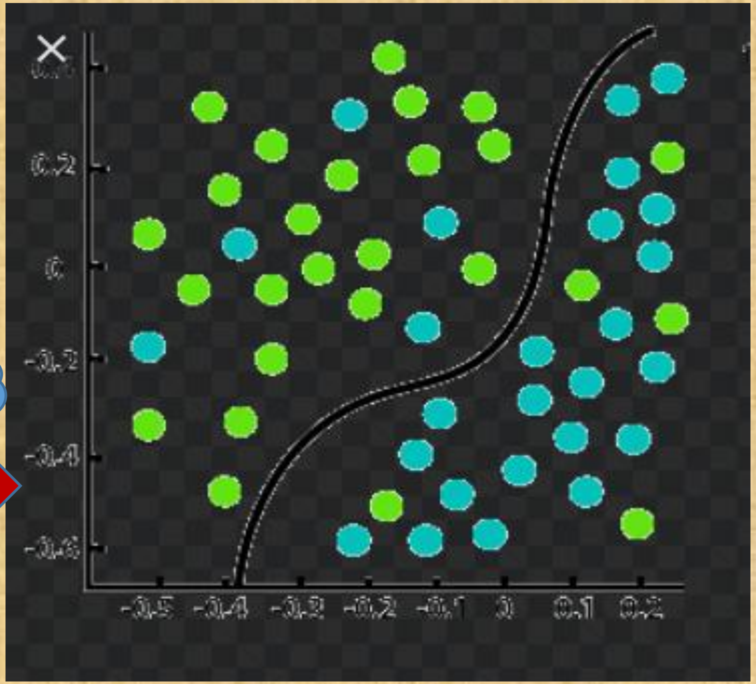
CLASSIFICATION



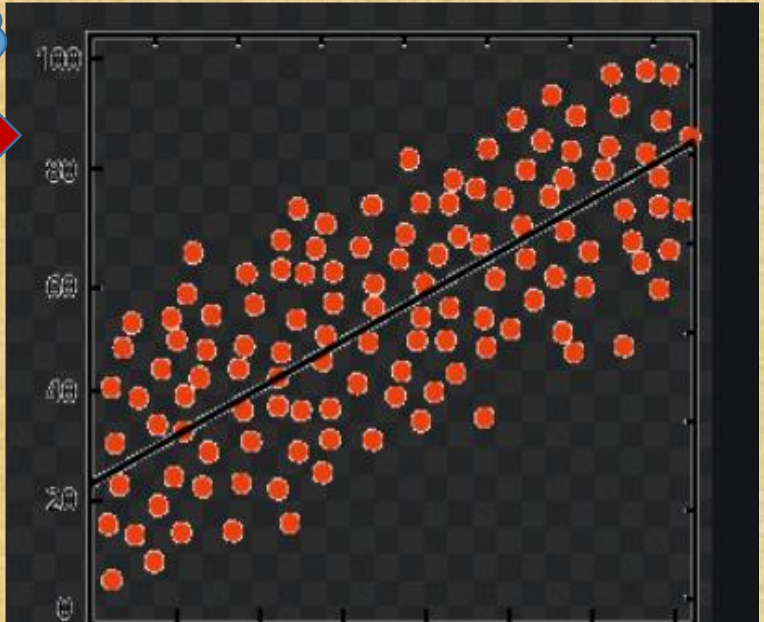




yes



no



DEFINITION.

- CLASSIFICATION is a datamining task of predicting the value of a categorical variable (target or class).
- Building a model based on one or more numerical and/or categorical variables (predictors, attributes or features)

Frequency Table

1. ZeroR
2. OneR
3. Naïve Bayesian
4. Decision Tree

Covariance Matrix

1. Linear Discriminant Analysis
2. Logistic Regression

Similarity functions

K Nearest Neighbours

Others

1. Artificial Neural Network
2. Support Vector Machine

LOGISTIC REGRESSION

Different ways of expressing probability

- Consider a two-outcome probability space, where:
  - $p(O_1) = p$
  - $p(O_2) = 1 - p = q$
- Can express probability of O_1 as:

	notation	range equivalents		
standard probability	p	0	0.5	1
odds	p / q	0	1	+ ∞
log odds (logit)	log(p / q)	- ∞	0	+ ∞

Log odds

- Numeric treatment of outcomes O_1 and O_2 is equivalent
  - If neither outcome is favored over the other, then log odds = 0.
  - If one outcome is favored with log odds = x, then other outcome is disfavored with log odds = $-x$.
- Especially useful in domains where relative probabilities can be miniscule
  - Example: multiple sequence alignment in computational biology

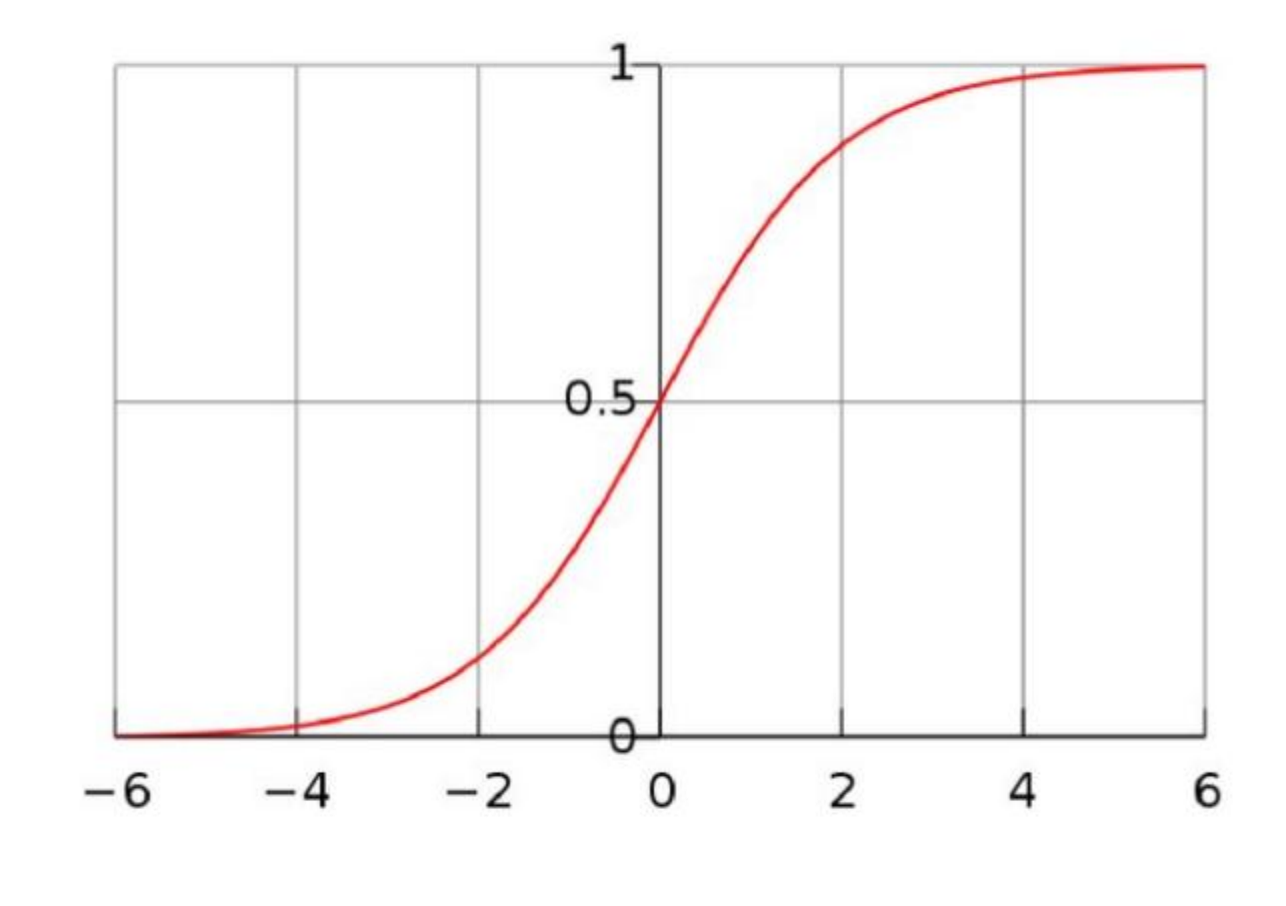
From probability to log odds (and back again)

$$z = \log\left(\frac{p}{1-p}\right)$$ logit function

$$\frac{p}{1-p} = e^z$$

$$p = \frac{e^z}{1+e^z} = \frac{1}{1+e^{-z}}$$ logistic function

STANDARD LOGISTIC FUNCTION



Scenario:

- A multidimensional feature space (features can be categorical or continuous).
- Outcome is discrete, not continuous.
  - We'll focus on case of two classes.
- It seems plausible that a linear decision boundary (hyperplane) will give good predictive accuracy.

USING A LOGISTIC REGRESSION MODEL

- Model consists of a vector **β** in *d*-dimensional feature space

- For a point **x** in feature space, project it onto **β** to convert it into a real number *z* in the range - ∞ to + ∞

$$z = \alpha + \boldsymbol{\beta} \cdot \mathbf{x} = \alpha + \beta_1 x_1 + \cdots + \beta_d x_d$$

- Map *z* to the range 0 to 1 using the logistic function

$$p = 1/(1 + e^{-z})$$

- Overall, logistic regression maps a point **x** in *d*-dimensional feature space to a value in the range 0 to 1

- Can interpret prediction from a logistic regression model as:
  - A probability of class membership
  - A class assignment, by applying threshold to probability
    - threshold represents decision boundary in feature space

TRAINING A LOGISTIC REGRESSION MODEL

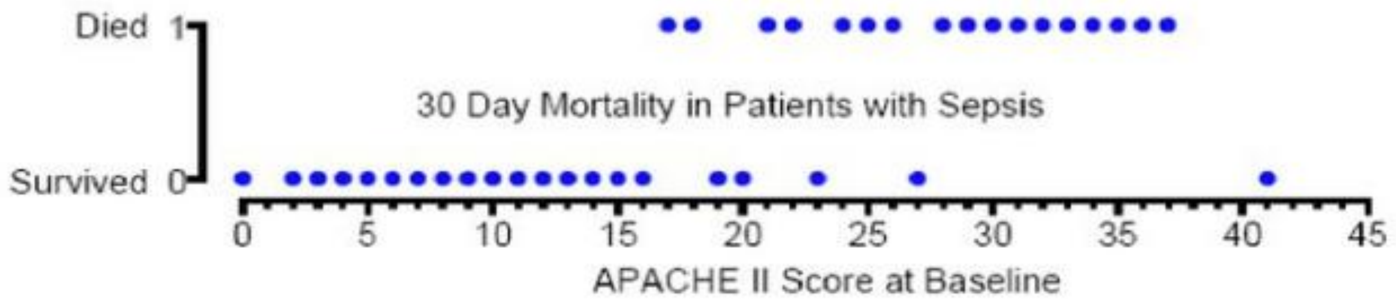
- Need to optimize β so the model gives the best possible reproduction of training set labels
  - Usually done by numerical approximation of maximum likelihood
  - On really large datasets, may use stochastic gradient descent

LOGISTIC REGRESSION IN ONE DIMENSION

WE WISH TO PREDICT DEATH FROM BASELINE APACHE II SCORE IN THESE PATIENTS.
LET *∏(x)* BE THE PROBABILITY THAT A PATIENT WITH SCORE *x* WILL DIE

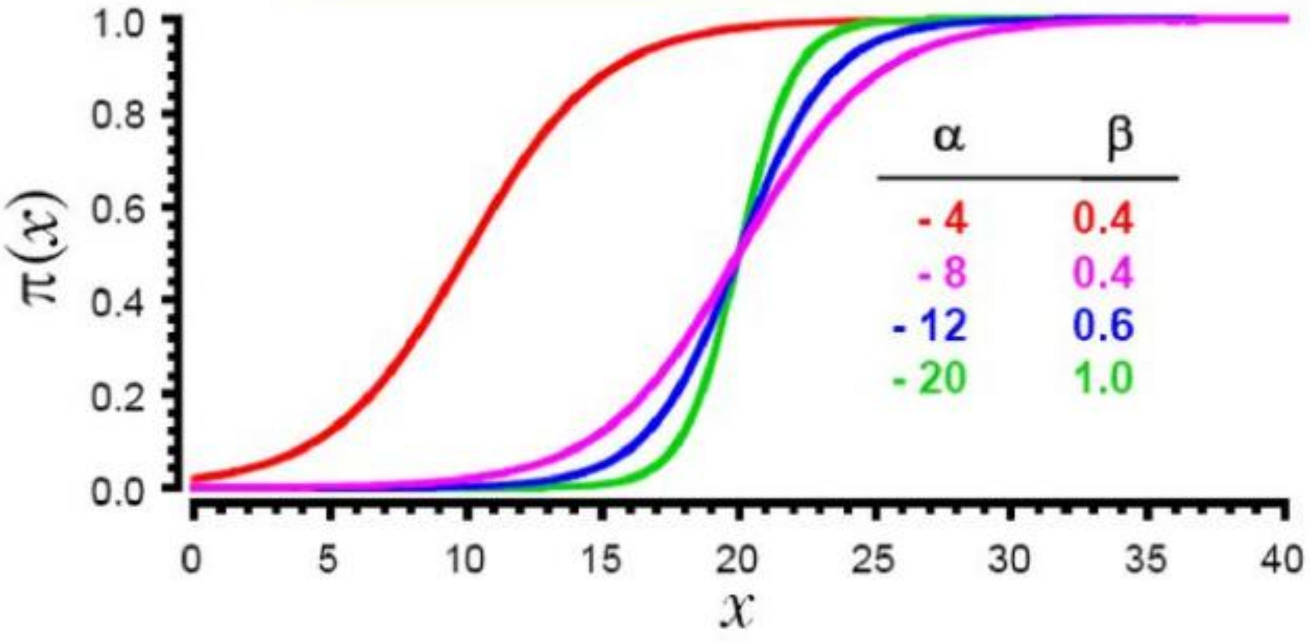
a) Example: APACHE II Score and Mortality in Sepsis

The following figure shows 30 day mortality in a sample of septic patients as a function of their baseline APACHE II Score. Patients are coded as 1 or 0 depending on whether they are dead or alive in 30 days, respectively.



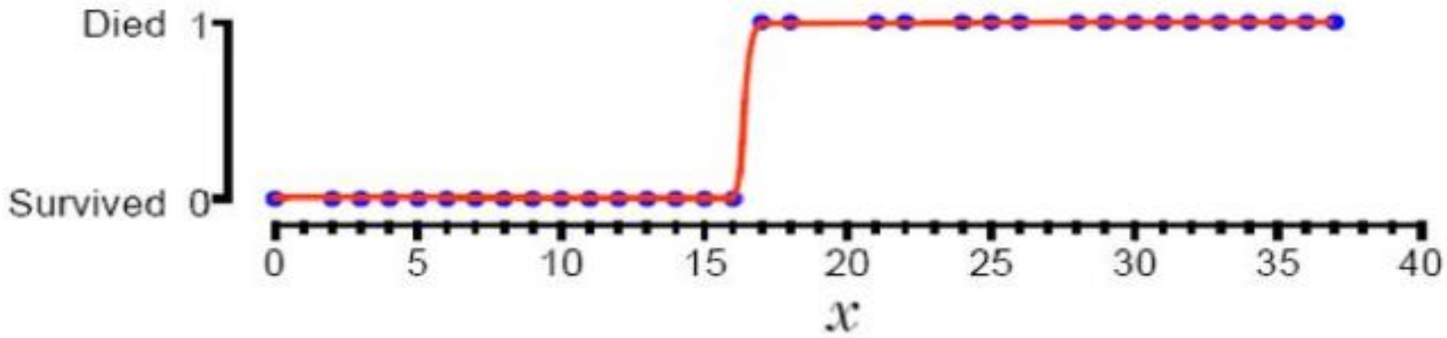
- Parameters control shape and location of sigmoid curve
  - α controls location of midpoint
  - β controls slope of rise

$$\pi(x) = \exp(\alpha + \beta x) / (1 + \exp(\alpha + \beta x))$$

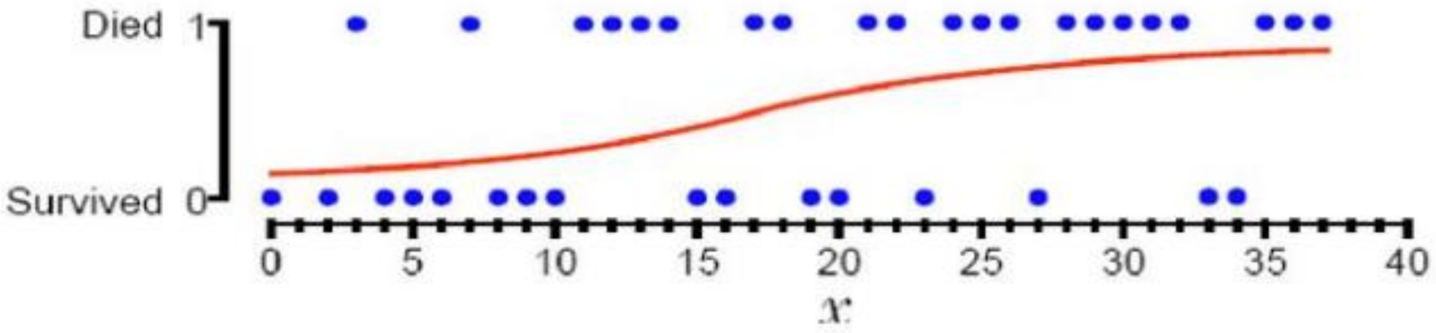


When $x = -\alpha / \beta$, $\alpha + \beta x = 0$ and hence $\pi(x) = 1/(1+1) = 0.5$

Data that has a sharp survival cut off point between patients who live or die should have a large value of β.



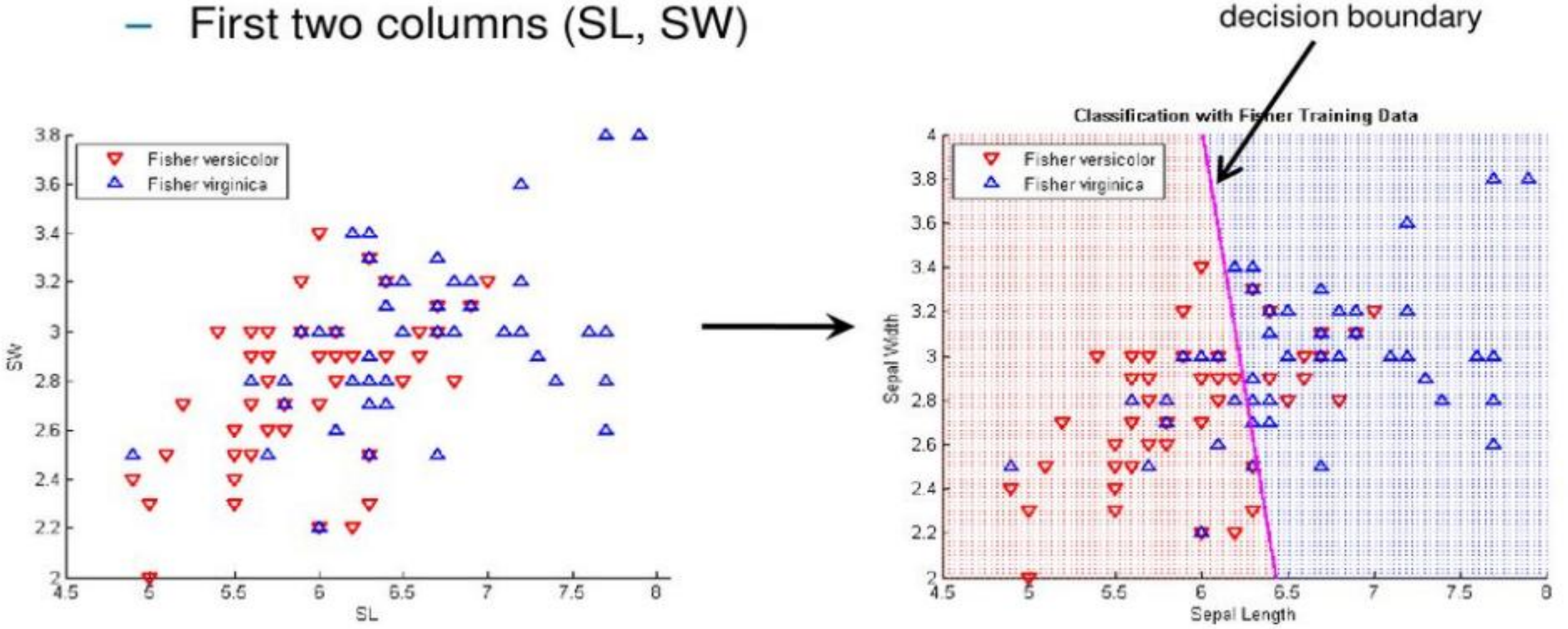
Data with a lengthy transition from survival to death should have a low value of β.



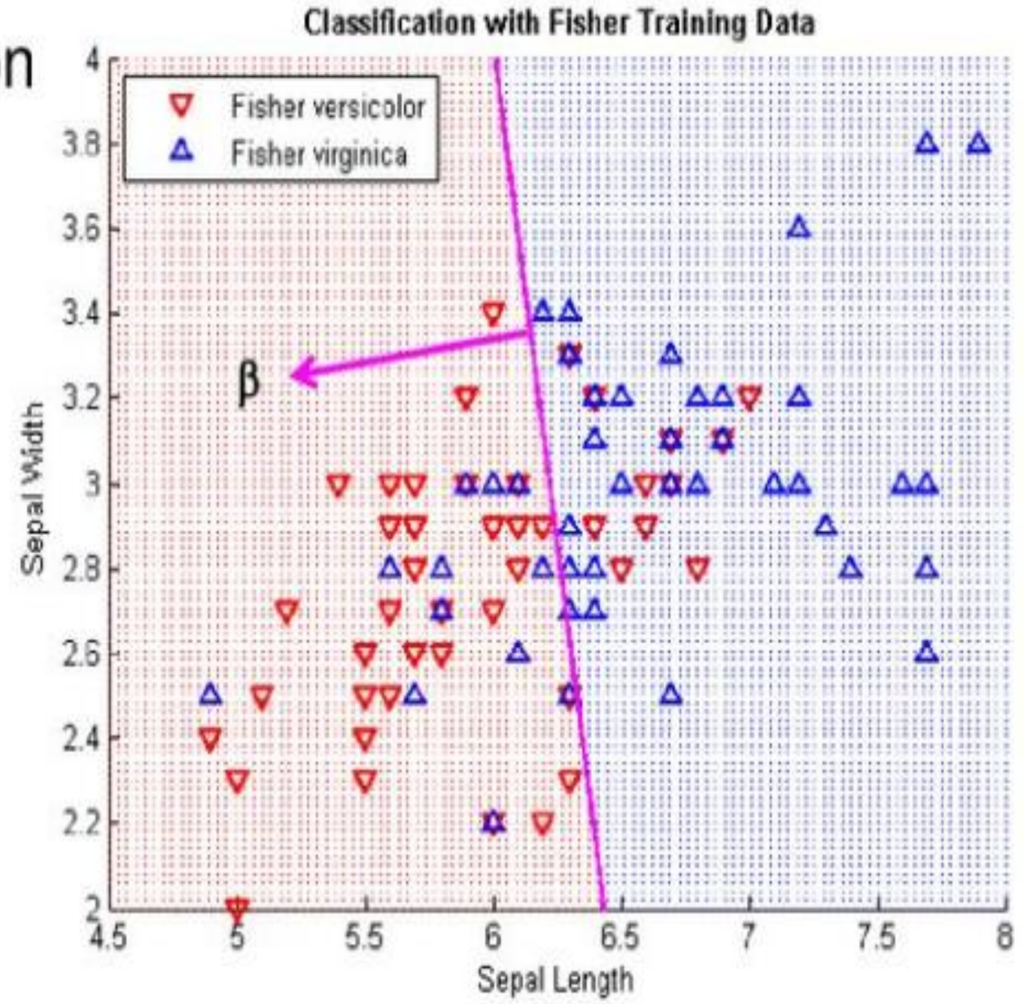
LOGISTIC REGRESSION IN TWO DIMENSION

Subset of Fisher iris dataset

- Two classes
- First two columns (SL, SW)



- α, β define location and orientation of decision boundary
  - $-\alpha$ is distance of decision boundary from origin
  - decision boundary is perpendicular to β
- magnitude of β defines gradient of probabilities between 0 and 1



HEART DISEASE & DIABETICS DATASET

- 13 attributes (see `heart.docx` for details)
  - 2 demographic (age, gender)
  - 11 clinical measures of cardiovascular status and performance
- 2 classes: absence (1) or presence (2) of heart disease
- 270 samples
- Dataset taken from UC Irvine Machine Learning Repository:

  http://archive.ics.uci.edu/ml/datasets/Statlog+(Heart)