MIDS W207
Applied Machine Learning

Week 5
Live Session Slides

Regression Analysis

Regression analysis is a set of statistical methods used for the estimation of relationships between a dependent variable and one or more independent variables.

It can be utilized to assess the strength of the relationship between variables and for modeling the future relationship between them.

Applications

1. Forecasting
2. Capital Asset Pricing Model (CAPM)
3. Comparing with competition
4. Identifying problems
5. Reliable Source

Regression Analysis

Regression analysis is a set of statistical methods used for the estimation of relationships between a dependent variable and one or more independent variables.

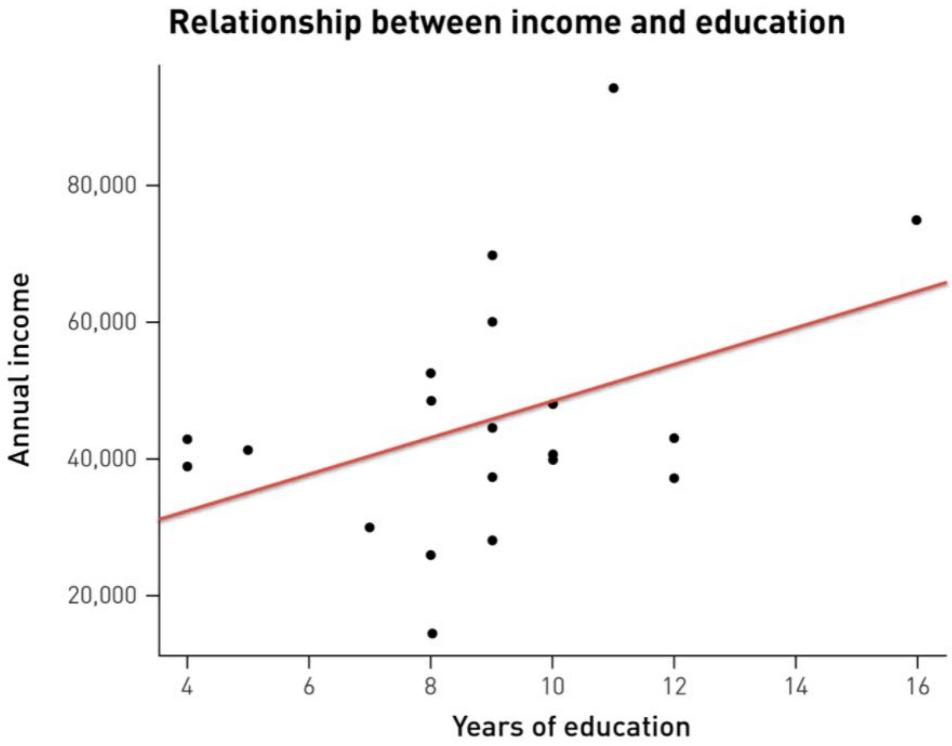
It can be utilized to assess the strength of the relationship between variables and for modeling the future relationship between them.

Applications

1. Forecasting
2. Capital Asset Pricing Model (CAPM)
3. Comparing with competition
4. Identifying problems
5. Reliable Source

$$\underset{\substack{\uparrow \\ \text{Dependent} \\ \text{Variable}}}{Y_i} = \overset{\substack{\text{Constant/Intercept} \\ \downarrow}}{\beta_0} + \underset{\substack{\uparrow \\ \text{Slope/Coefficient}}}{\beta_1} \overset{\substack{\text{Independent} \\ \text{Variable} \\ \downarrow}}{X_i}$$

Regression Analysis: Example



Regression Analysis: Notations

Subscript Notation

$$y_i = \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + \varepsilon_i = x_i^T \beta + \varepsilon_i$$

$$i = 1, \ldots, n$$

Matrix Notation

$$y = X\beta + \varepsilon$$

$$\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} \mathbf{x}_1^{\mathrm{T}} \\ \mathbf{x}_2^{\mathrm{T}} \\ \vdots \\ \mathbf{x}_n^{\mathrm{T}} \end{pmatrix} = \begin{pmatrix} x_{11} & \cdots & x_{1p} \\ x_{21} & \cdots & x_{2p} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{np} \end{pmatrix}, \quad \boldsymbol{\beta} = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{pmatrix}, \quad \boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

Regression Analysis: Types

Linear Regression assumes linear relationship between dependent and independent variables.

$$y_i = a + bx_i + \varepsilon_i$$

Multiple Regression is a linear regression with multiple independent variables.

$$y_i = \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + \varepsilon_i = \mathrm{x}_i^{\mathrm{T}} \beta + \varepsilon_i, \; i = 1, \ldots, n$$

Logistic Regression considers binary dependent variables.

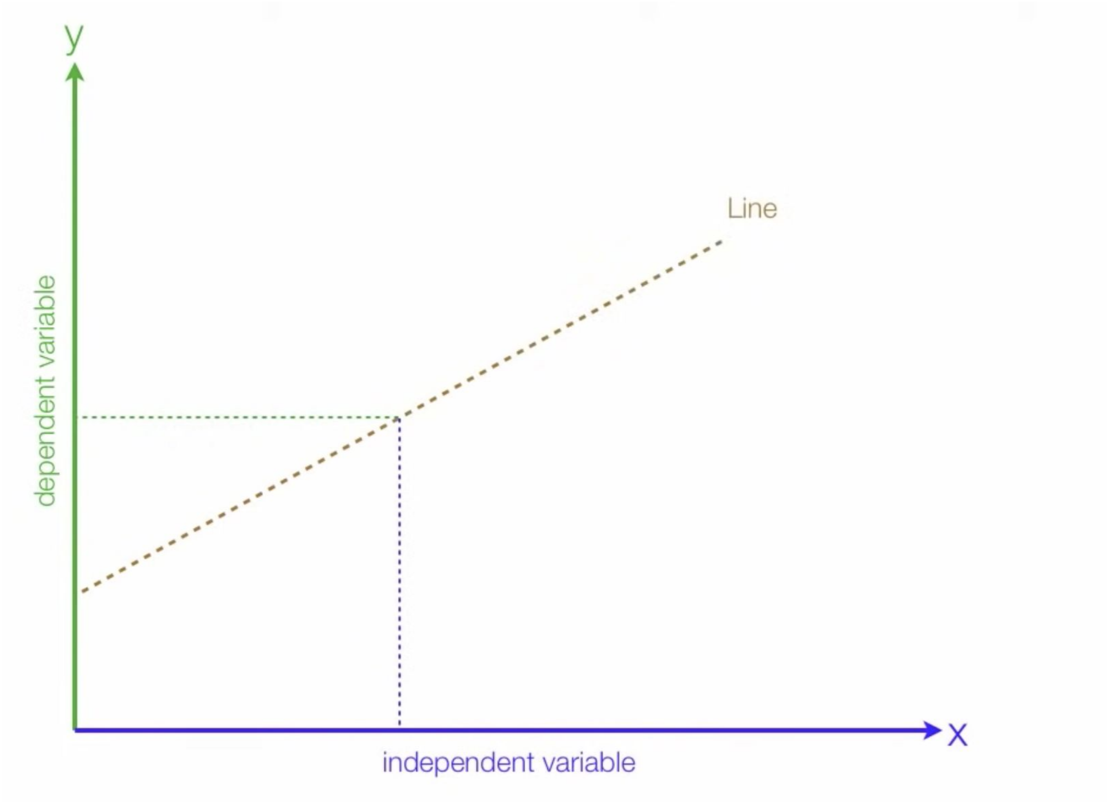
$$\operatorname{logit}\left(p_i\right) = \ln\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 x_{1,i} + \cdots + \beta_M x_{m,i}$$

Polynomial Regression considers relationship between the independent variable x and the dependent variable y is modelled as an nth degree polynomial in x.

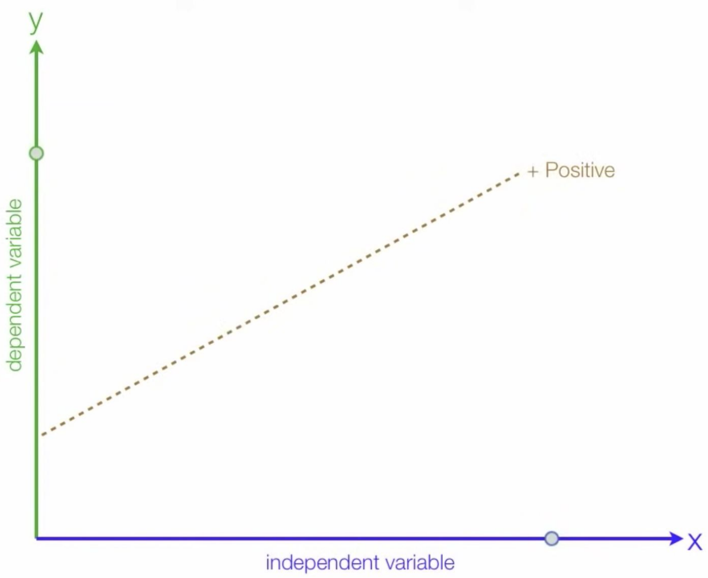
$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \cdots + \beta_n x^n + \varepsilon.$$

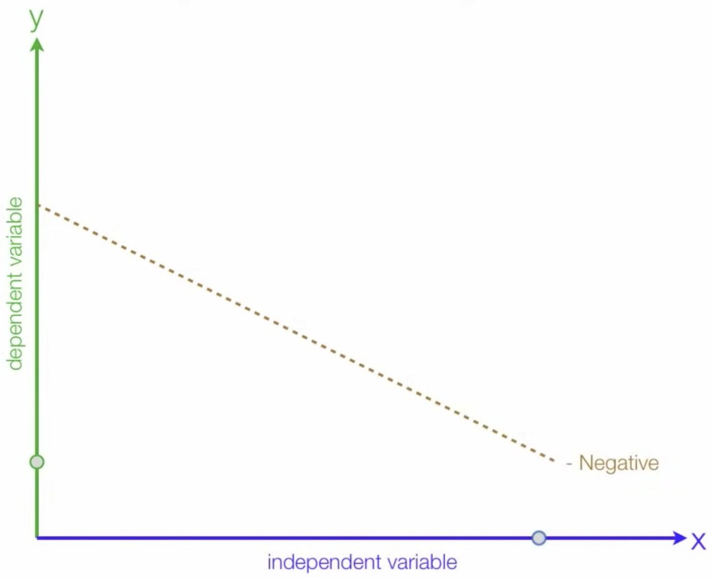
Kernel Regression is a non-parametric technique to estimate the conditional expectation of a random variable. The objective is to find a non-linear relation between a pair of random variables X and Y.

Regression Example

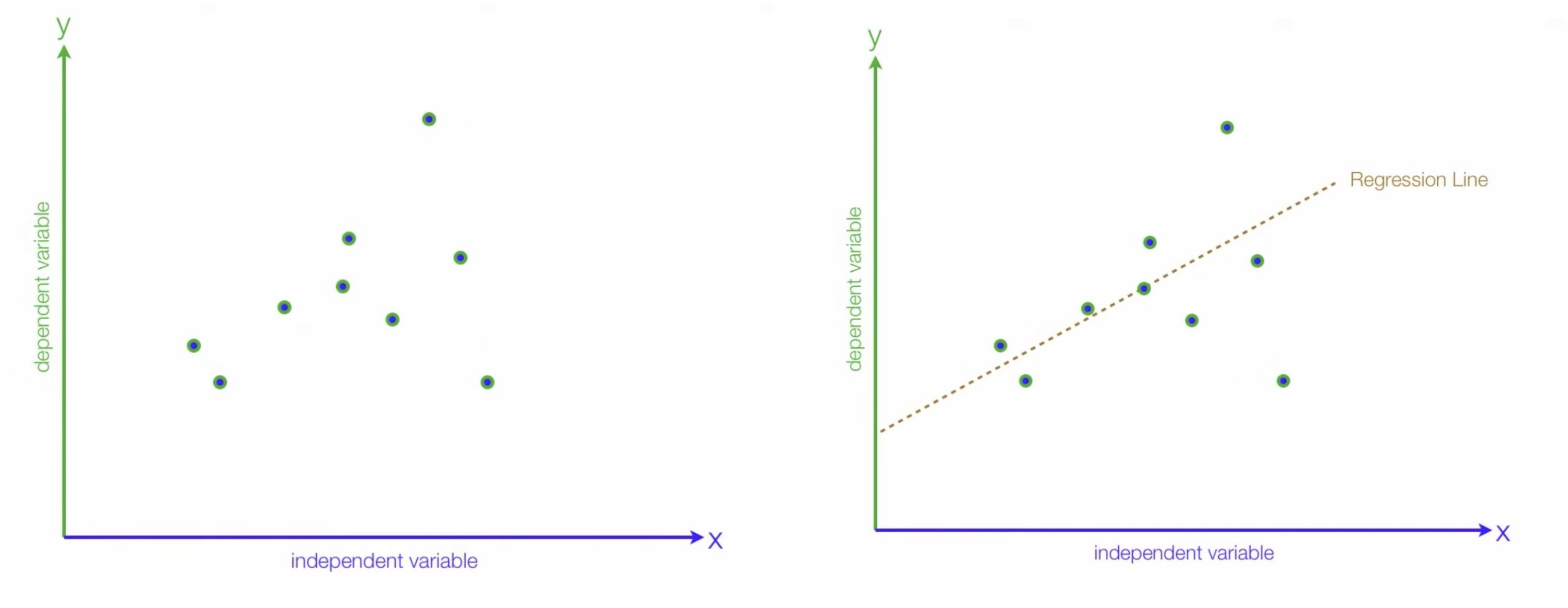


Regression Example

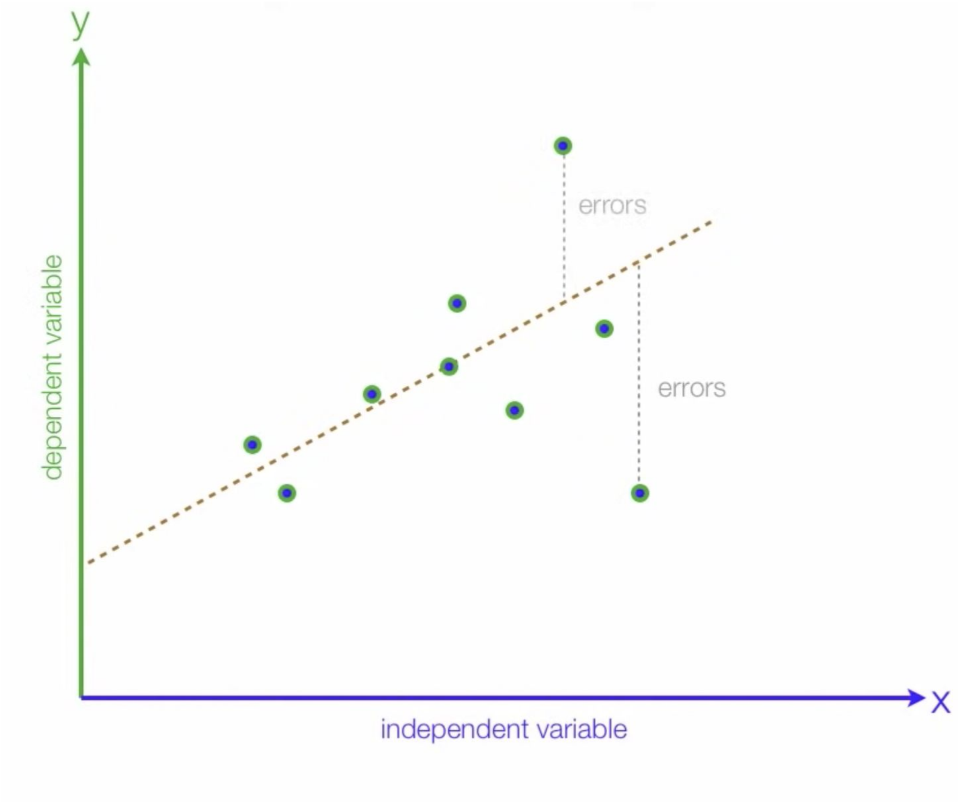




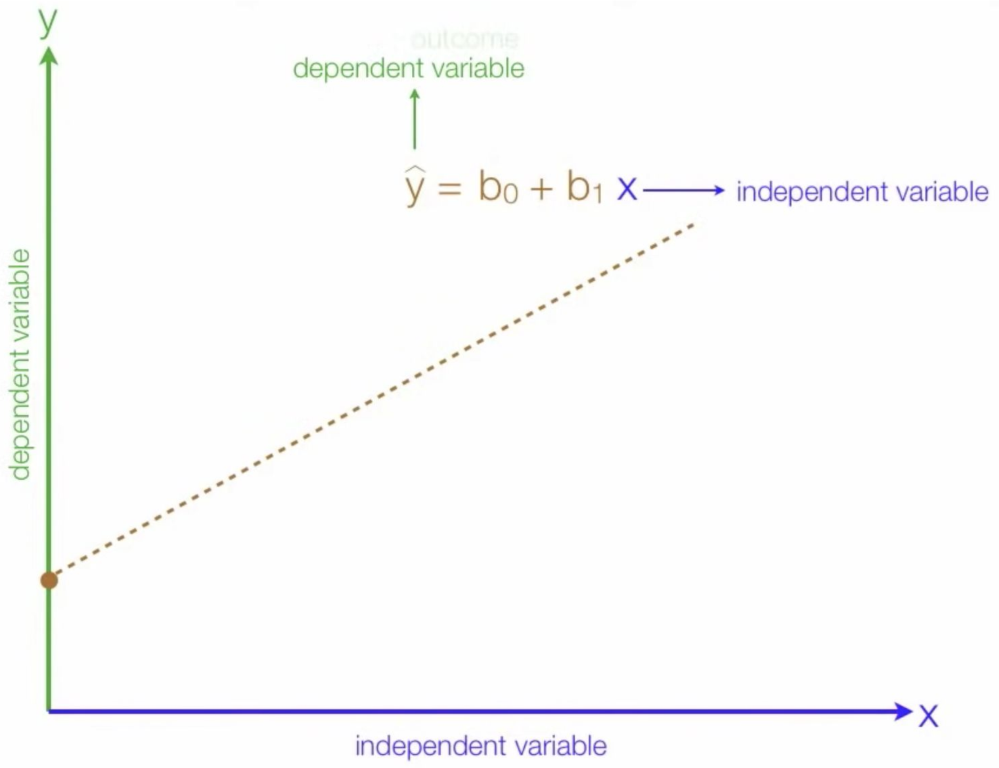
Regression Example



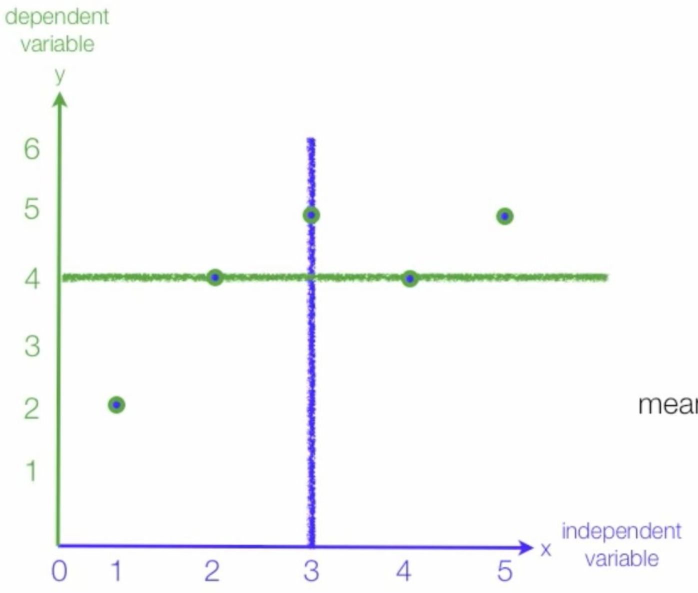
Regression Example



Regression Example

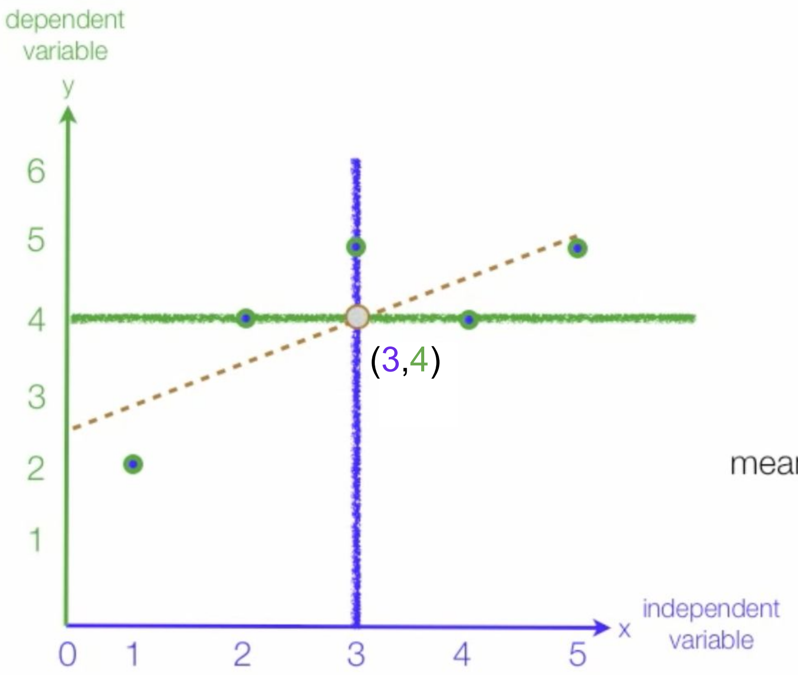


Regression Example



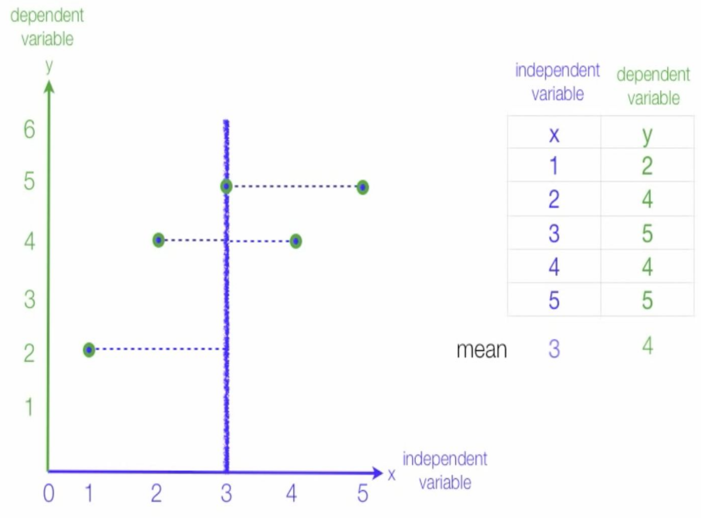
	independent variable	dependent variable
	x	y
	1	2
	2	4
	3	5
	4	4
	5	5
mean	3	4

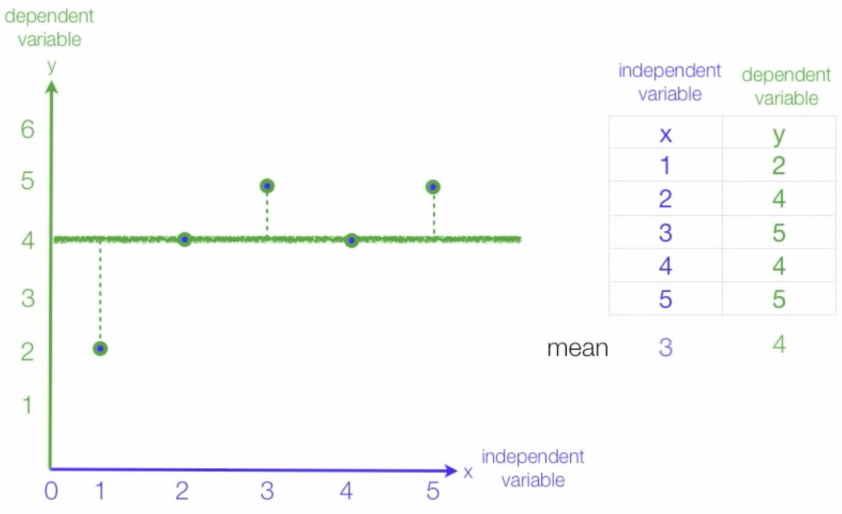
Regression Example



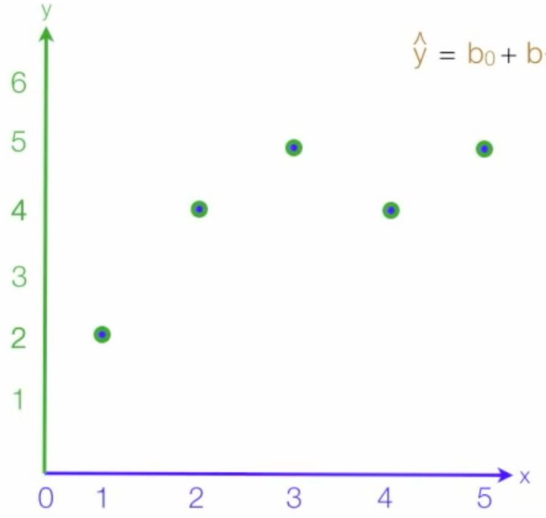
	independent variable	dependent variable
	x	y
	1	2
	2	4
	3	5
	4	4
	5	5
mean	3	4

Regression Example



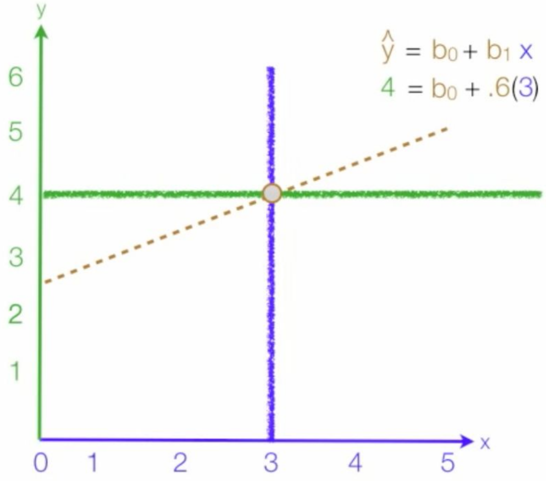


Regression Example



	x	y	$x - \bar{x}$	$y - \bar{y}$	$(x - \bar{x})^2$	$(x - \bar{x})(y - \bar{y})$
	1	2	-2	-2	4	4
	2	4	-1	0	1	0
	3	5	0	1	0	0
	4	4	1	0	1	0
	5	5	2	1	4	2
mean	3	4				

Regression Example



$\hat{y} = b_0 + b_1 x$

$4 = b_0 + .6(3)$

$b_0 = 2.2$

$b_1 = .6$

$\hat{y} = 2.2 + .6x$

	x	y	$x - \bar{x}$	$y - \bar{y}$	$(x - \bar{x})^2$	$(x - \bar{x})(y - \bar{y})$
	1	2	-2	-2	4	4
	2	4	-1	0	1	0
	3	5	0	1	0	0
	4	4	1	0	1	0
	5	5	2	1	4	2
					10	6
mean	3	4				

$4 = b_0 + .6(3)$

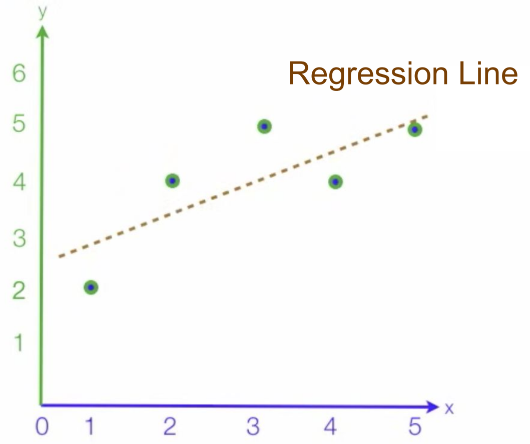
$4 = b_0 + \cancel{1.8}$

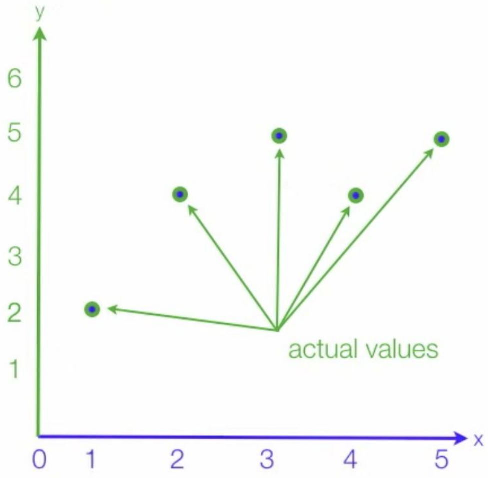
$-1.8 \quad\quad -\cancel{1.8}$

$2.2 = b_0$

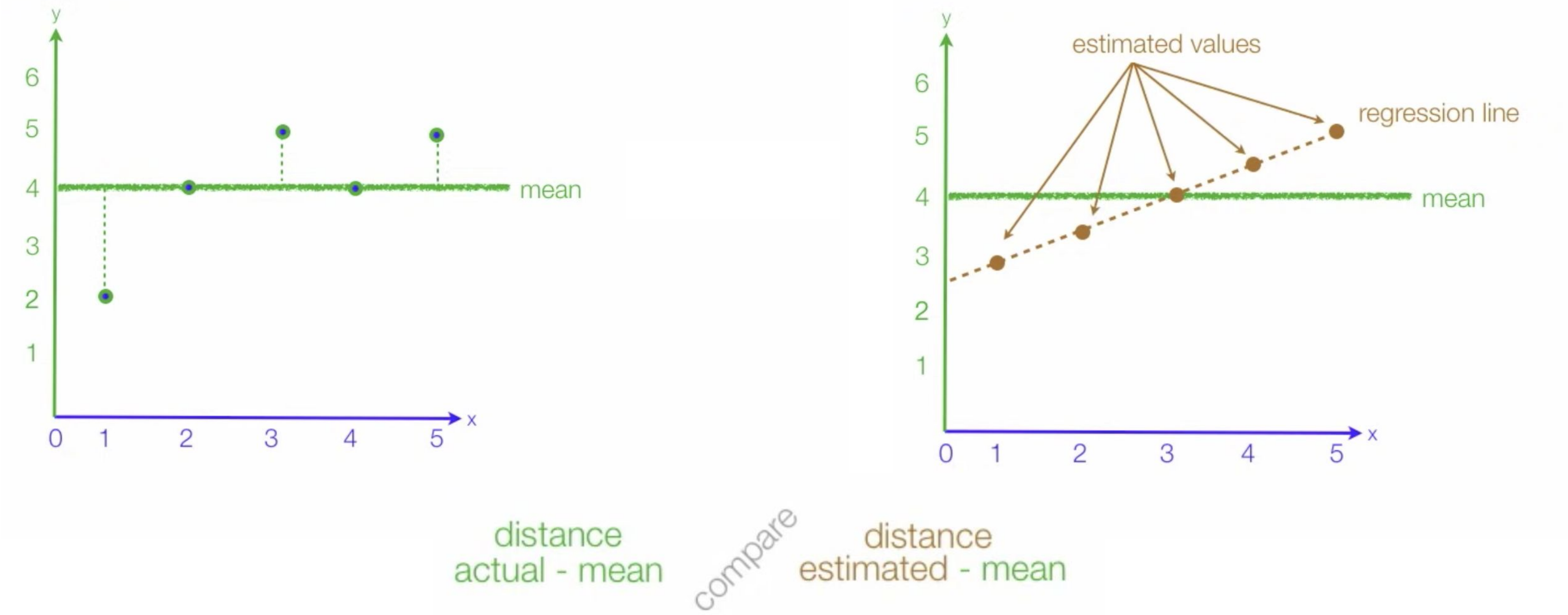
$$b_1 = \frac{6}{10} = .6 = \frac{\Sigma(x - \bar{x})(y - \bar{y})}{\Sigma(x - \bar{x})^2}$$

Regression Example: R-Squared

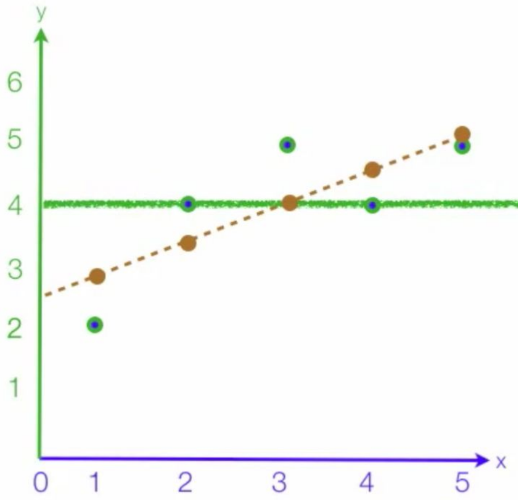




Regression Example: R-Squared



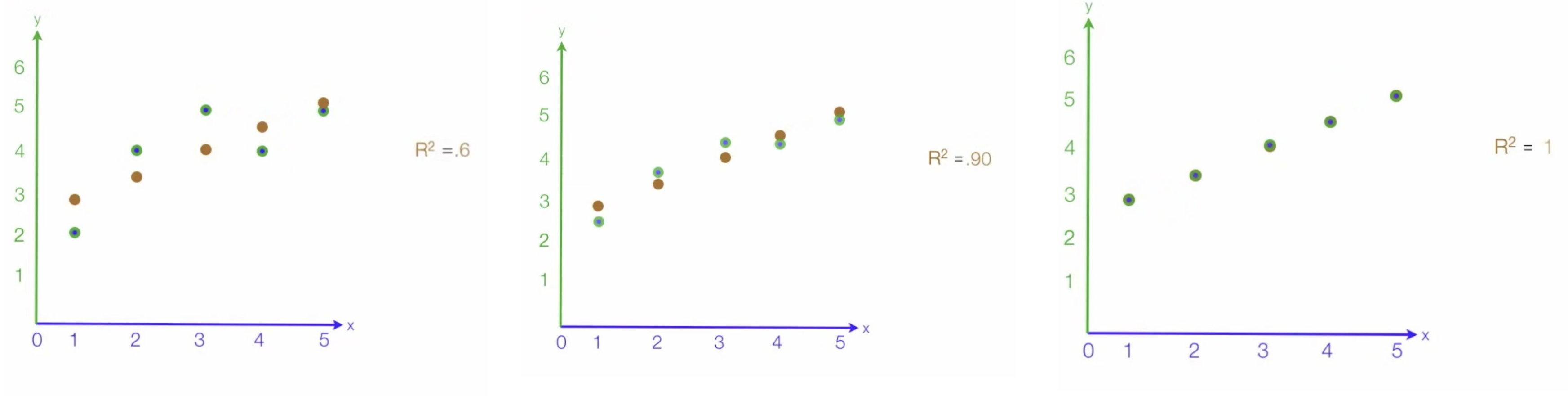
Regression Example: R-Squared



x	y	$y - \bar{y}$	$(y - \bar{y})^2$	$\hat{y}$	$\hat{y} - \bar{y}$	$(\hat{y} - \bar{y})^2$
1	2	-2	4	2.8	-1.2	1.44
2	4	0	0	3.4	-.6	.36
3	5	1	1	4	0	0
4	4	0	0	4.6	.6	.36
5	5	1	1	5.2	1.2	1.44
			6			3.6
mean	4					

$$R^2 = \frac{3.6}{6} = .6 = \frac{\Sigma(\hat{y} - \bar{y})^2}{\Sigma(y - \bar{y})^2}$$

Regression Example: R-Squared



Regression: Error Minimization (OLS)

$$Y_i = \alpha + \beta X_i + \varepsilon_i$$

$$\varepsilon_i = Y_i - \hat{Y}_i = Y_i - \alpha - \beta X_i$$

$$\min_{\alpha,\beta} \sum_{i=1}^{N} \varepsilon_i^2 = \sum_{i=1}^{N} (Y_i - \alpha - \beta X_i)^2$$

$$\hat{\beta} = \left(\mathbf{X}^\top \mathbf{X}\right)^{-1} \mathbf{X}^\top \mathbf{y}$$

$\hat{\beta}$ = ordinary least squares estimator

X = matrix regressor variable X

$\top$ = matrix transpose

y = vector of the value of the response variable

Regression: R

Regression in *R*

- **lm( )** fits a linear regression (**l**inear **m**odel).

```
results <- lm(outcome ~ predictor1 + ... + predictorP)
```

- **summary( )** or **display( )** shows output.

```
> summary(lm(income~education+foreign, data=sls))

Call:
lm(formula = income ~ education + foreign, data = sls)

Residuals:
      Min        1Q    Median        3Q       Max
   -28203    -11191      -431      9788     43579

Coefficients:
             Estimate Std. Error  t value   Pr(>|t|)
Intercept       19379      15710    1.234     0.2342
education        2866       1578    1.817     0.0869
foreignyes       2483       9906    0.251     0.8051

Residual standard error: 17550 on 17 degrees of freedom
Multiple R-squared: 0.1739
Adjusted R-squared: 0.0767
F-statistic: 1.789 on 2 and 17 DF
p-value: 0.1972
```

Linear Regression: Paradigms

Statistician's (inferential)	Computer science (predictive)
What is the effect of an additional year of schooling on wages?	What do we expect an individual with 14 years of education to earn?
How much revenue per day was lost when the server went down?	How much revenue do we expect today, if 12 servers are running?

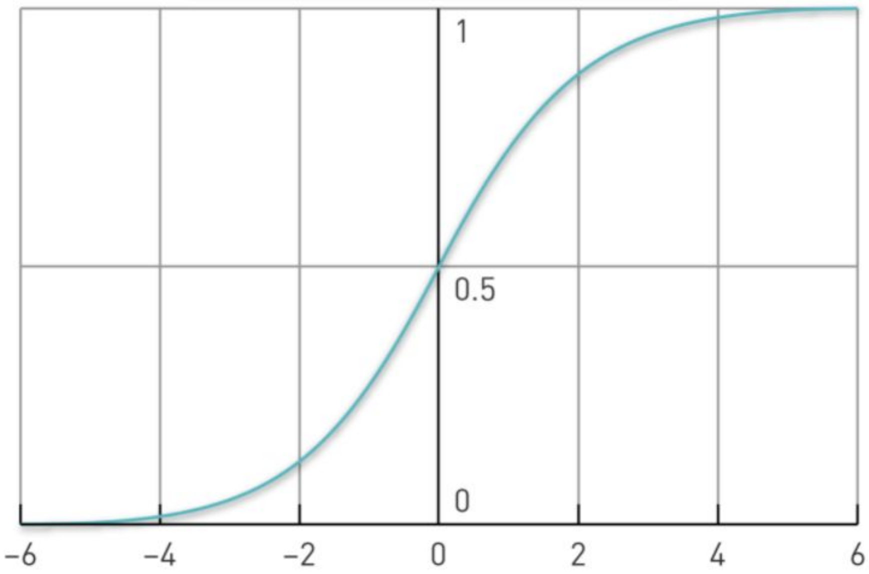
Inferential Model

- Given data with known values
- Fit a model
- Determine relationships in training data

Predictive Model

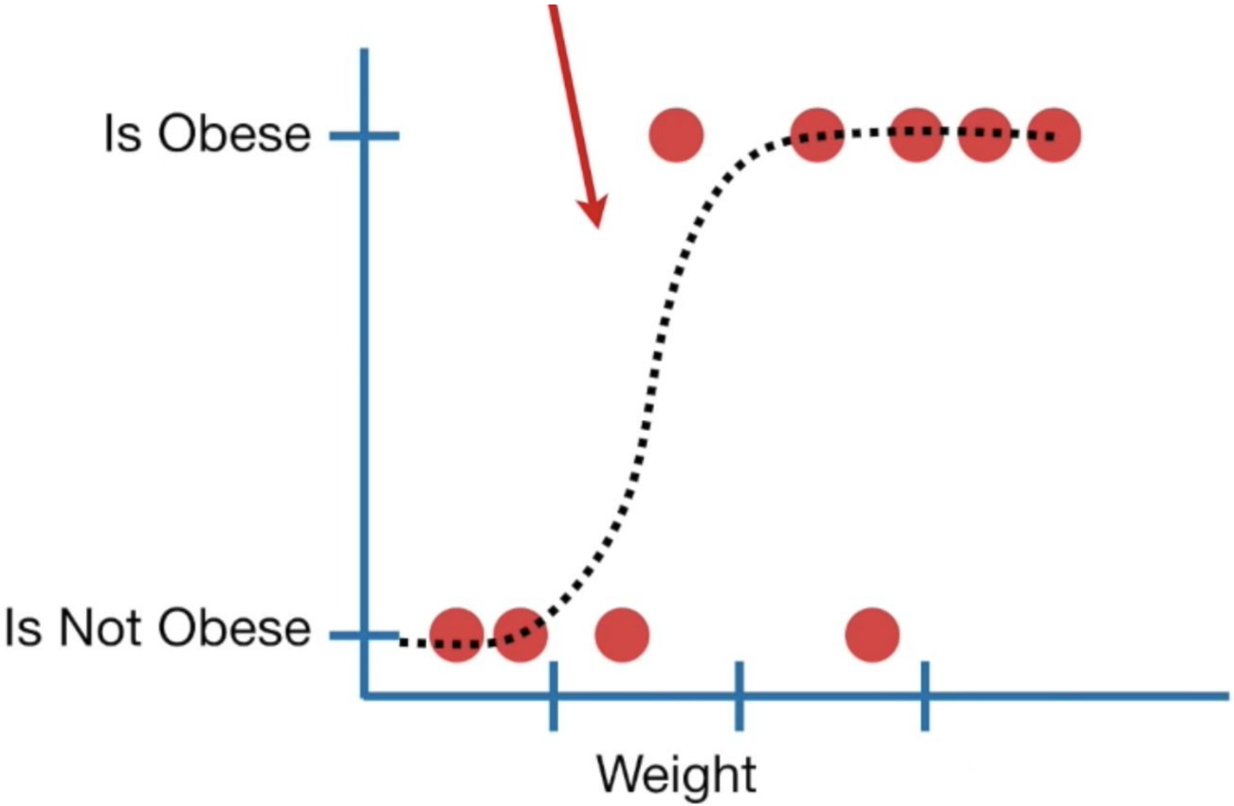
- Given outcomes of training data
- Learn relationship between dependent and independent variables
- Predict unknown values in test data

Logistic Regression

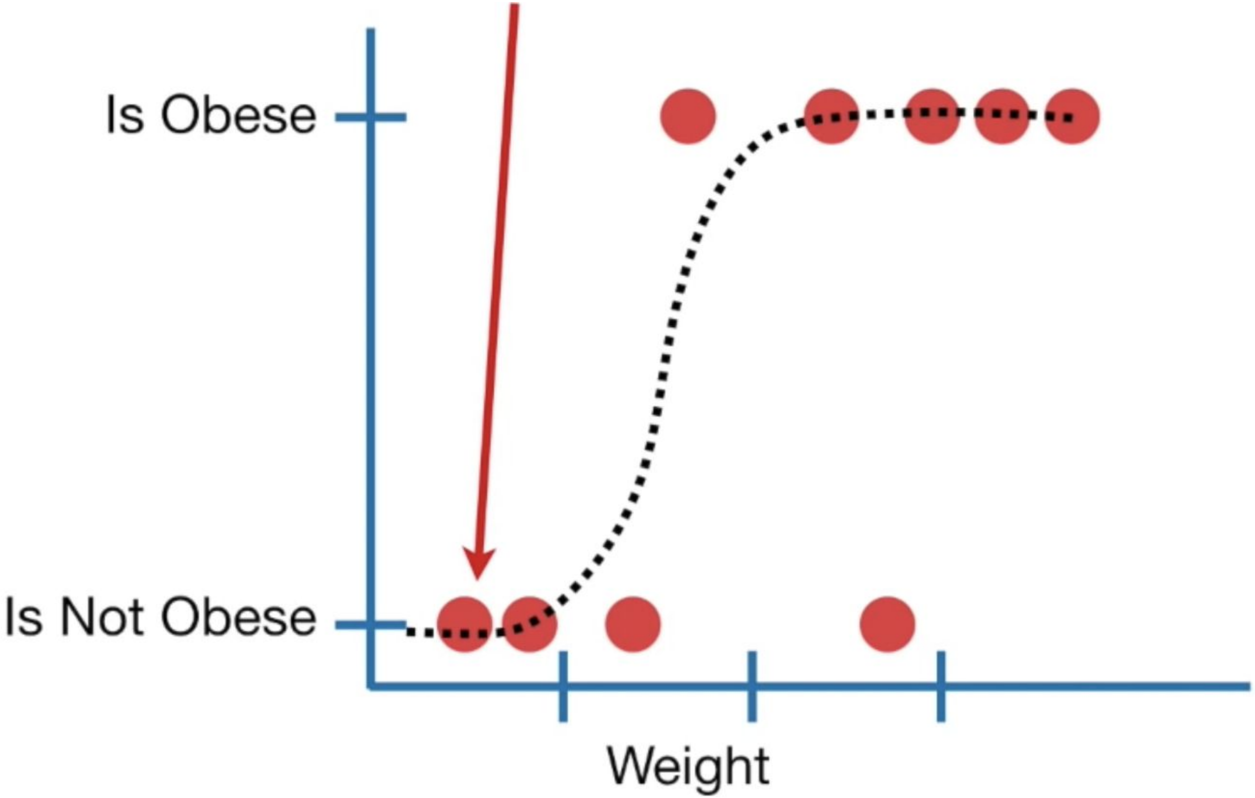


$$P = \frac{e^{\alpha+\beta X}}{1+e^{\alpha+\beta X}}$$

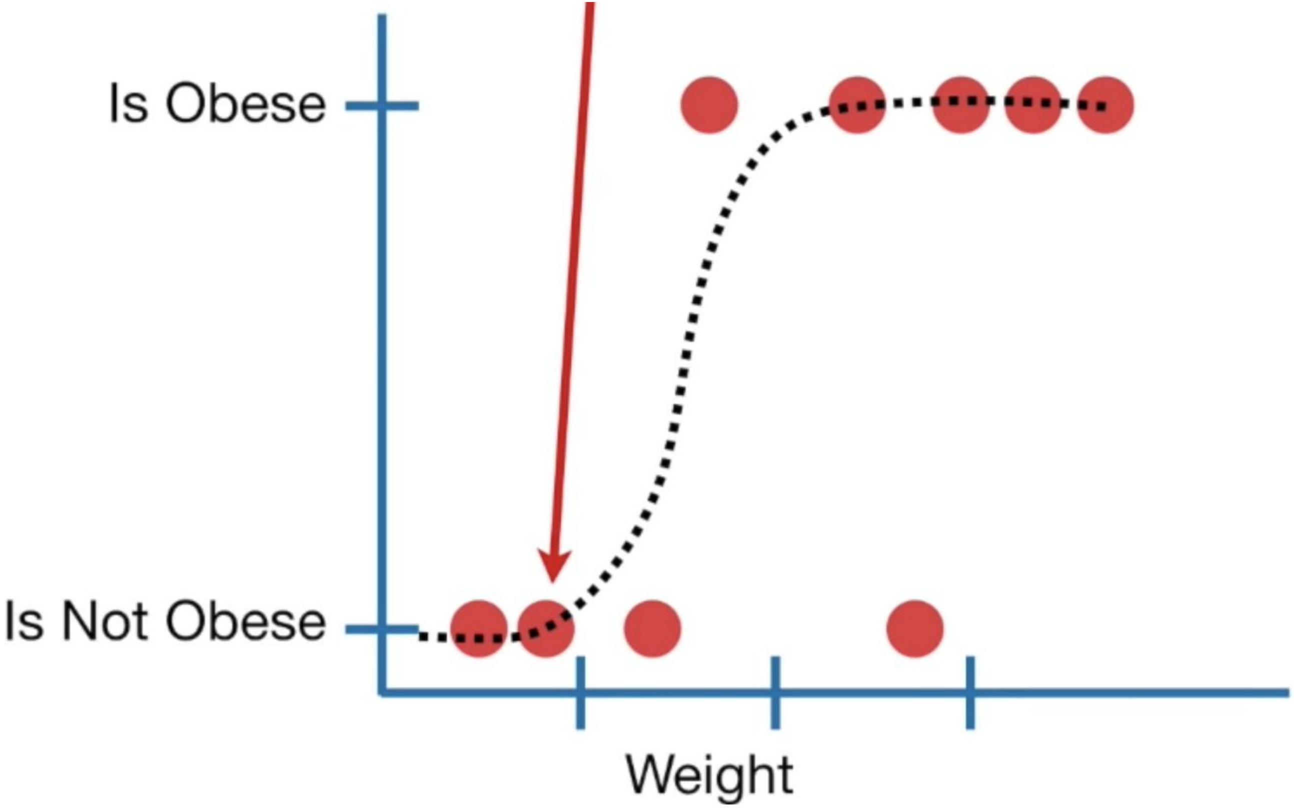
Logistic Regression: Maximum Likelihood Estimate



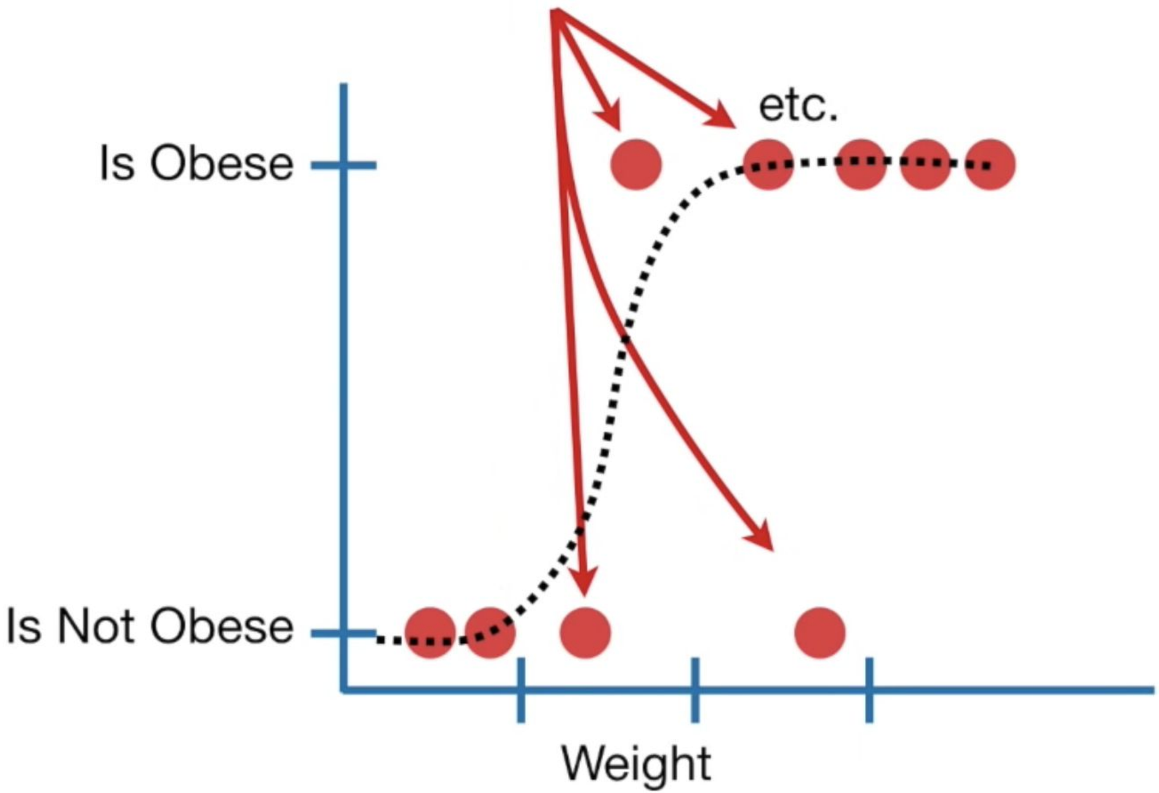
Logistic Regression: Maximum Likelihood Estimate



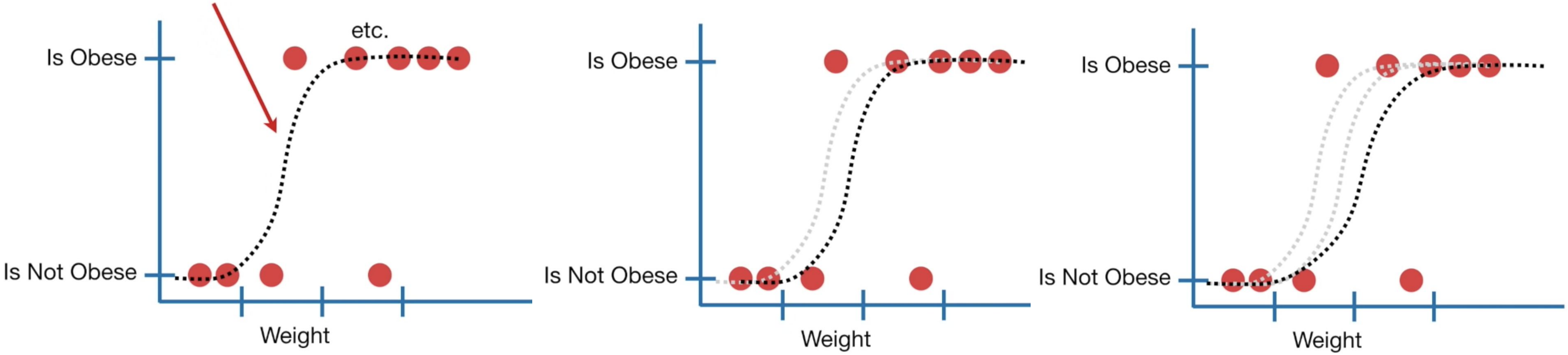
Logistic Regression: Maximum Likelihood Estimate



Logistic Regression: Maximum Likelihood Estimate



Logistic Regression: Maximum Likelihood Estimate



Regression: Thresholds in Supervised Learning

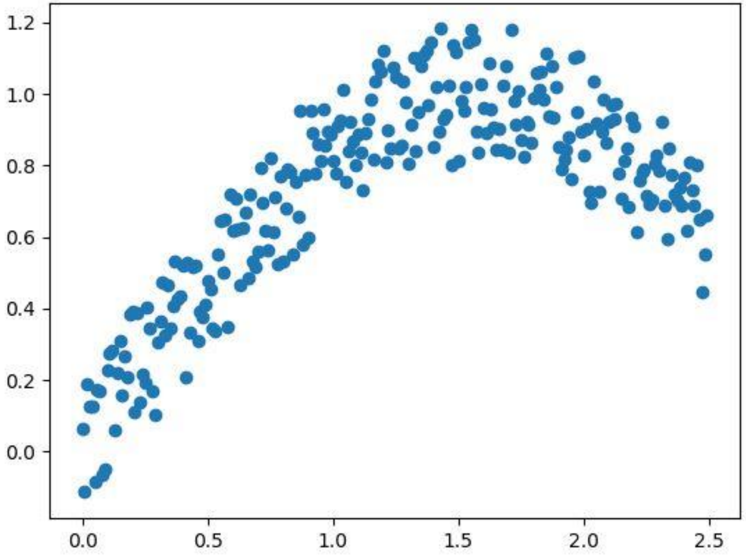
Confusion Matrix

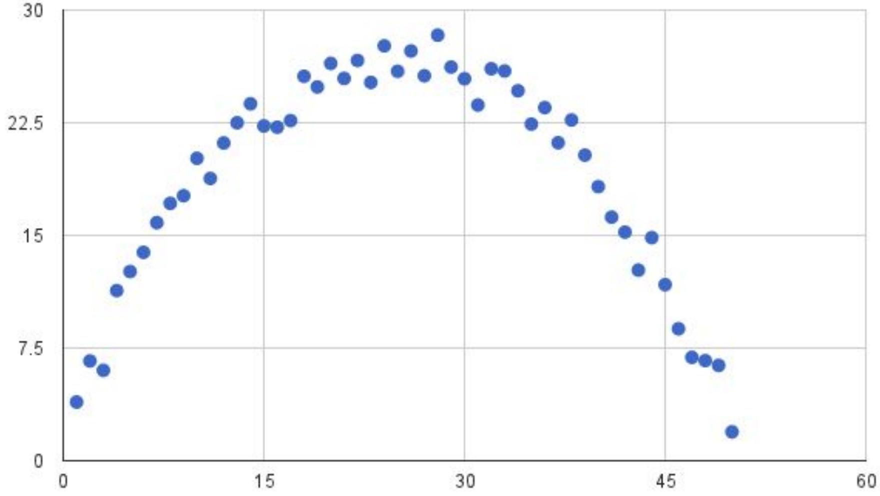
		Actual Value	
		Yes (1)	No (0)
Predicted Value	Yes (1)	TP	FP
	No (0)	FN	TN

TP= True Positive
FP= False Positive
FN= False Negative
TN= True Negative

- If you have supervised data, you will want to maximize an objective function.
  - **Precision**: $TP \div (TP + FP)$ % positives correctly identifed
  - **Recall**: $TP \div (TP + FN)$ % existing positives identified
  - **Optimal point** on ROC (precision/recall) curve
  - **Accuracy**: $(TP + TN) \div (TP + TN + FP + FN)$
  - **F-test**: $2 \cdot (P \cdot R) \div (P + R)$

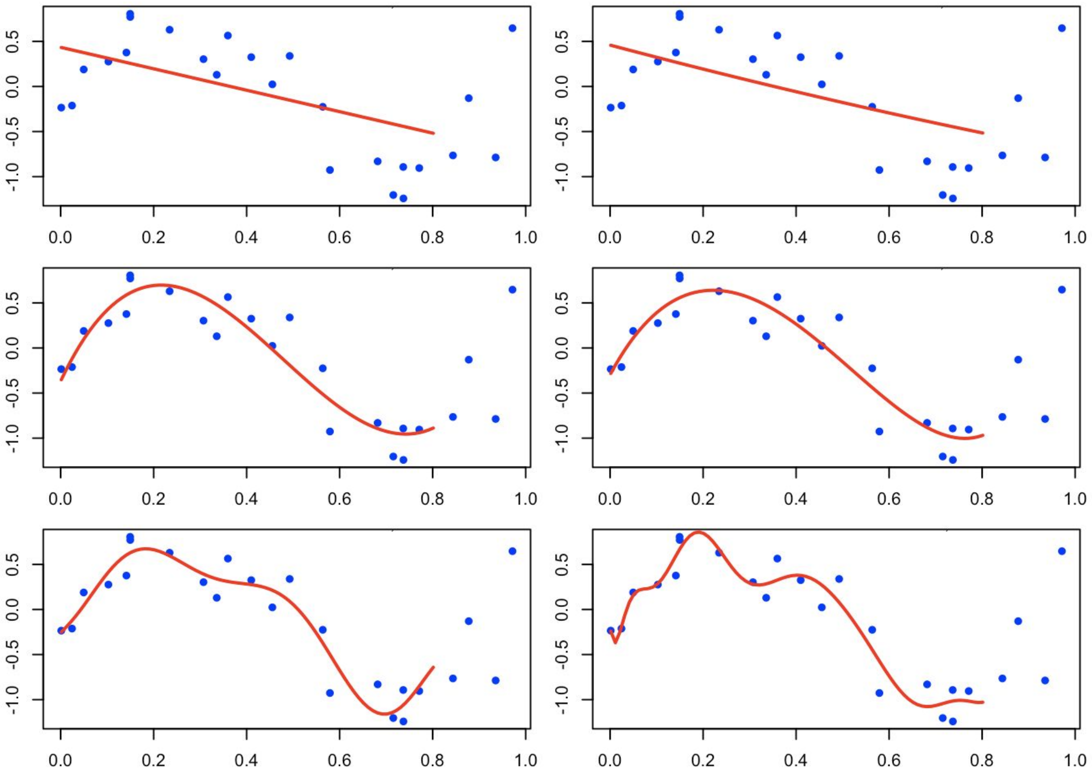
Regression: Extensions: Polynomial Regression





$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \cdots + \beta_n x^n + \varepsilon.$$

Regression: Extensions: Polynomial Regression



Regression: Extensions: Kernel Regression

Kernel Regression/Classification Algorithm

1. Given training data $D = \{\mathbf{x}_i, y_i\}$, Kernel function $K(\cdot,\cdot)$ and input $\mathbf{x}$
   - (regression) if $y \in \mathbf{R}$, return weighted average:

$$\hat{y}(\mathbf{x}) = \frac{\sum_{i=1}^{n} K(\mathbf{x},\mathbf{x}_i) y_i}{\sum_{i=1}^{n} K(\mathbf{x},\mathbf{x}_i)}$$

   - (classification) if $y \in \pm 1$, return weighted majority:

$$\hat{y}(\mathbf{x}) = sign(\sum_{i=1}^{n} K(\mathbf{x},\mathbf{x}_i) y_i)$$