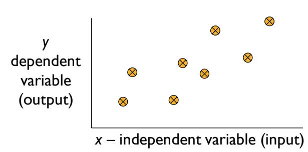
REGRESSION

DR. SHAILESH S

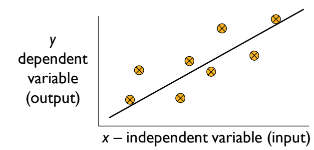
REGRESSION

- For classification the output(s) is nominal
- In regression the output is continuous
  - Function Approximation
- Many models could be used – Simplest is linear regression
  - Fit data with the best hyper-plane which "goes through" the points



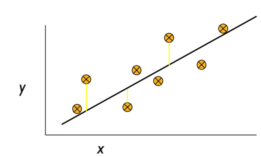
REGRESSION

- For classification the output(s) is nominal
- In regression the output is continuous
  - Function Approximation
- Many models could be used – Simplest is linear regression
  - Fit data with the best hyper-plane which "goes through" the points



REGRESSION

- For classification the output(s) is nominal
- In regression the output is continuous
  - Function Approximation
- Many models could be used – Simplest is linear regression
  - Fit data with the best hyper-plane which "goes through" the points
  - For each point the differences between the predicted point and the actual observation is the *residue*



SIMPLE LR

Weight (lbs)	Height (inches)
140	60
155	62
159	67
179	70
192	71
200	72
212	75

$$Height = \beta_0 + \beta_1\, Weight$$

$$Y = \beta_0 + \beta_1 X$$

MULTIPLE LR

Price	Advertisement	Promotion	Unit Sales
$8.75	$50.04	$61.13	73959
$8.99	$50.74	$60.19	71544
$7.50	$50.14	$59.16	78587
$7.25	$50.27	$60.38	80364
$7.40	$51.25	$59.71	78771
$8.50	$50.65	$59.88	71986
$8.40	$50.87	$60.14	74885
$7.90	$50.15	$60.08	73345
$7.25	$48.24	$59.90	76659
$8.70	$50.19	$59.68	71880
$8.40	$51.11	$59.83	73598
$8.10	$51.49	$59.77	74893
$8.40	$50.10	$59.29	69003
$7.40	$49.24	$60.40	78542
$8.00	$50.04	$59.89	72543
$8.30	$49.46	$60.06	74247
$8.10	$51.62	$60.51	76253
$8.20	$49.78	$58.93	72582
$8.99	$48.60	$60.09	69022
$7.99	$49.00	$61.00	76200
$8.50	$48.00	$59.00	69701
$7.90	$54.00	$59.50	77005
$7.99	$48.70	$58.00	70987
$8.25	$50.00	$60.50	75643

X Y

$$sales = \beta_0 + \beta_1\, price + \beta_2\, promo + \beta_3\, ad$$

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \ldots + \hat{\beta}_p x_p$$

$$\mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} \quad \mathbf{X} = \begin{bmatrix} 1 & X_{11} & \cdots & X_{1p} \\ 1 & X_{21} & & X_{2p} \\ \vdots & \vdots & & \vdots \\ 1 & X_{n1} & & X_{np} \end{bmatrix} \quad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}$$

$$Y = X \bullet \beta$$

HOW DO WE "LEARN" PARAMETERS

- For the 2-*d* problem (line) there are coefficients for the bias and the independent variable (*y*-intercept and slope)

$$Y = \beta_0 + \beta_1 X$$

- To find the values for the coefficients which minimize the objective function

$$\beta_1 = \frac{n\sum xy - \sum x \sum y}{n\sum x^2 - \left(\sum x\right)^2} \qquad \beta_0 = \frac{\sum y - \beta_1 \sum x}{n}$$

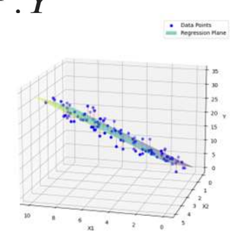
HOW DO WE "LEARN" PARAMETERS

- Multivariate Linear Regression

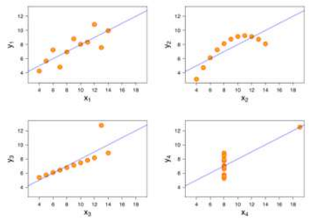
$$\hat{\beta} = (X^T . X)^{-1} . X^T . Y$$

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad \mathbf{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix}$$

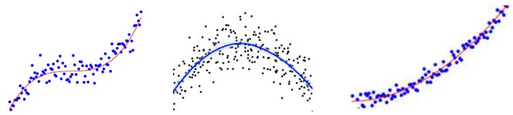
$$\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}, \quad \epsilon = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}$$



NON LINEAR REGRESSION



What lines "really" best fit each case? – different approaches



EVALUATION METRICS