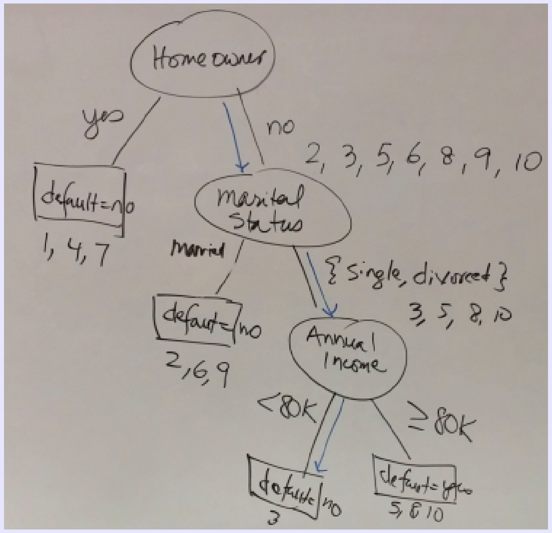
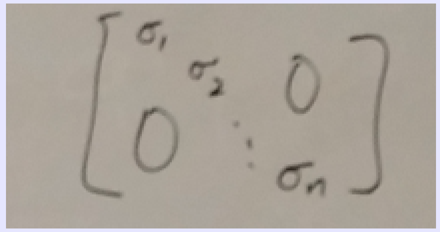
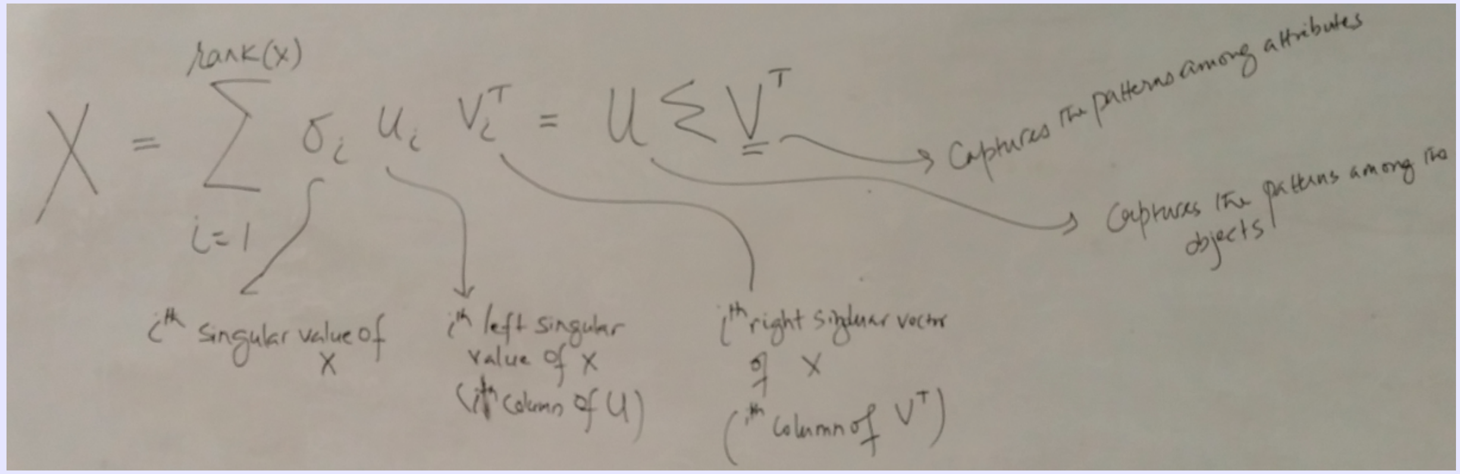
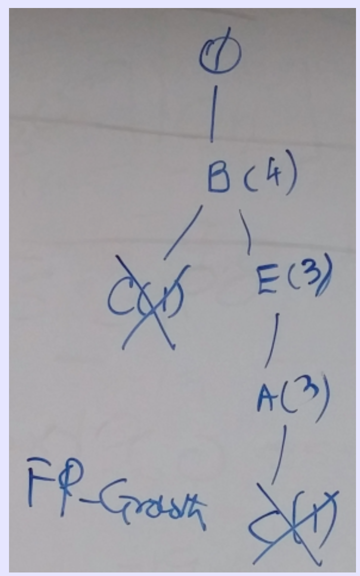
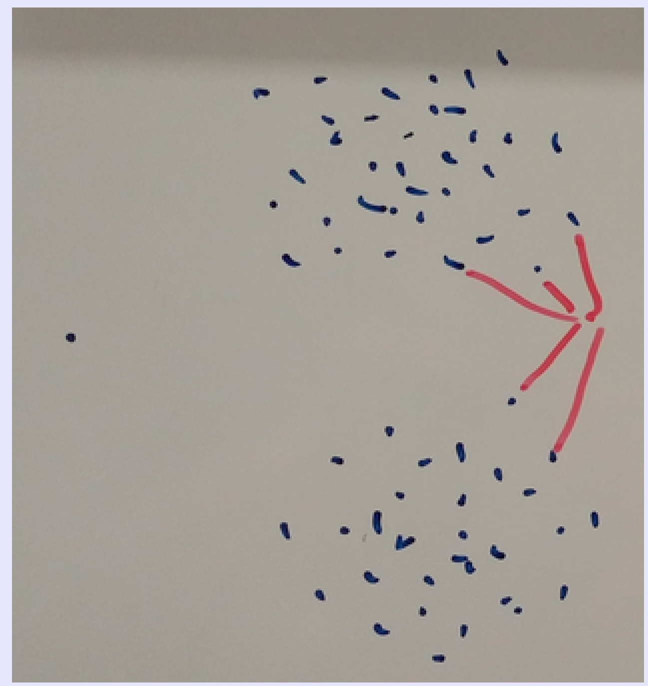










CS 422: Data Mining

Vijay K. Gurbani, Ph.D.,
Illinois Institute of Technology

Lecture: **Linear regression**

Linear regression: Theory

- A statistical process for estimating the relationship among variables.
  - The response variable (dependent variable, *Y*)
  - The predictor(s) variable(s) (independent variable(s), *X*)
- Used widely for **predicting** and **forecasting**.
- We will study method of least squares estimation technique.
  - Many other estimation techniques.

Linear regression: Theory

- The simple linear regression equation: $Y \approx \beta_0 + \beta_1 X$

Linear regression: Theory

- The simple linear regression equation: $Y \approx \beta_0 + \beta_1 X$

$$Y \approx \underset{\text{Intercept}}{\beta_0} + \underset{\text{Slope}}{\beta_1} X$$

- β_0 and β_1 are also called model *coefficients* or *parameters,* or *weights*.
- We use the dataset we have to produce *estimates* of $\hat{\beta}_0$ $\hat{\beta}_1$ for prediction on new values of *X*: $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$

Linear regression: Theory

- Because β$_0$ and β$_1$ are *estimates*, there will some error in the observed response (y_i) and predicted response ($\hat{y}_i$).
  - This error (hopefully small) is called the *residual*, (ε), or $\epsilon_i = y_i - \hat{y}_i = (\beta_0 + \beta_1 x_i) - (\hat{\beta}_0 + \hat{\beta}_1 x_i)\ \forall i$
  - $\hat{\beta}_0$ $\hat{\beta}_1$ are chosen to minimize the sum of residuals (RSS, or residual sum of squares):

$$RSS = \epsilon_1^2 + \epsilon_2^2 + ... + \epsilon_n^2$$

  - In other words, we set up the following optimization problem: $\min_{\hat{\beta}_0, \hat{\beta}_1} \sum_{i=1}^{N} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2.$

Linear regression: Theory

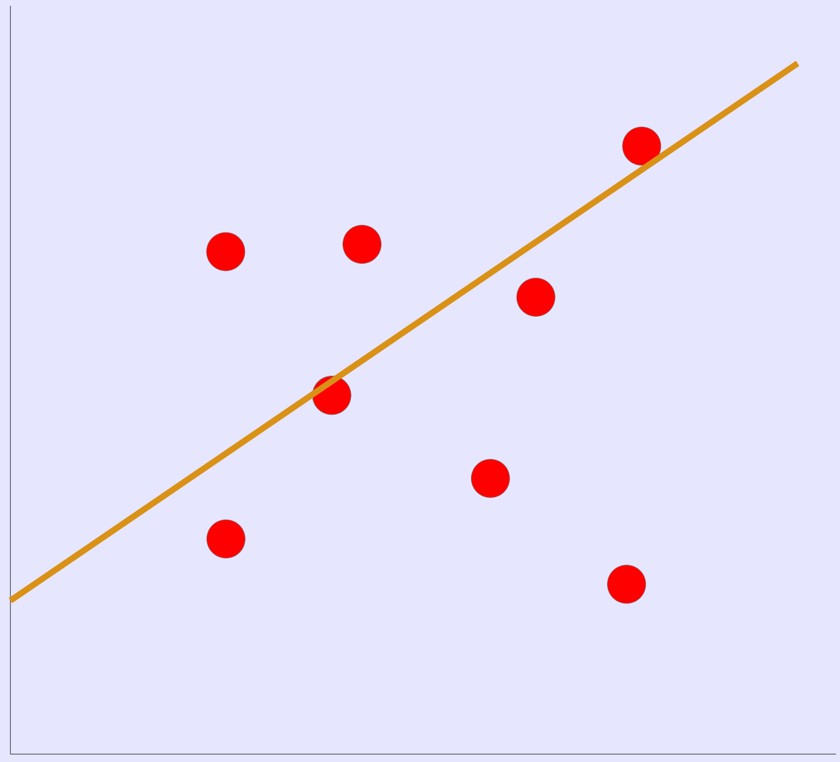
- Turns out that $\hat{\beta}_0$ $\hat{\beta}_1$ can be derived analytically from the RSS using some calculus.

$$\hat{\beta}_1 = \frac{\sum(y_i - \bar{y})(x_i - \bar{x})}{\sum(x_i - \bar{x})^2}$$
$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x}$$

- Once you have $\hat{\beta}_0$ $\hat{\beta}_1$, prediction follows the linear equation just derived: $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$ for values of *X*.
- In reality, the relationship is shown below, where ε is the catch-all (error) fo r what is missed by the model: $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i + \epsilon$

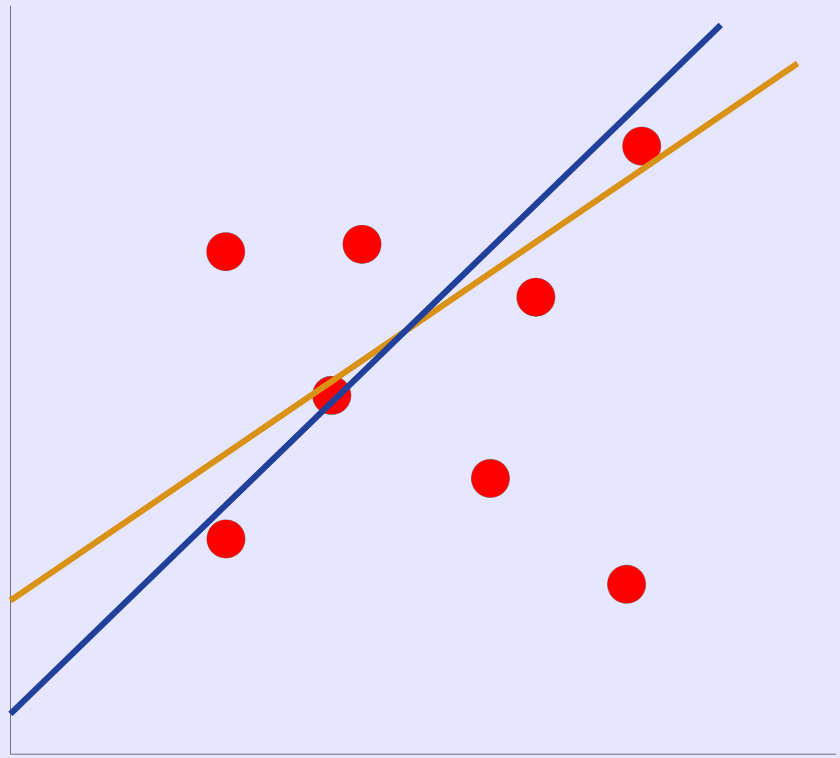
Linear regression: Theory

- Geometric interpretation



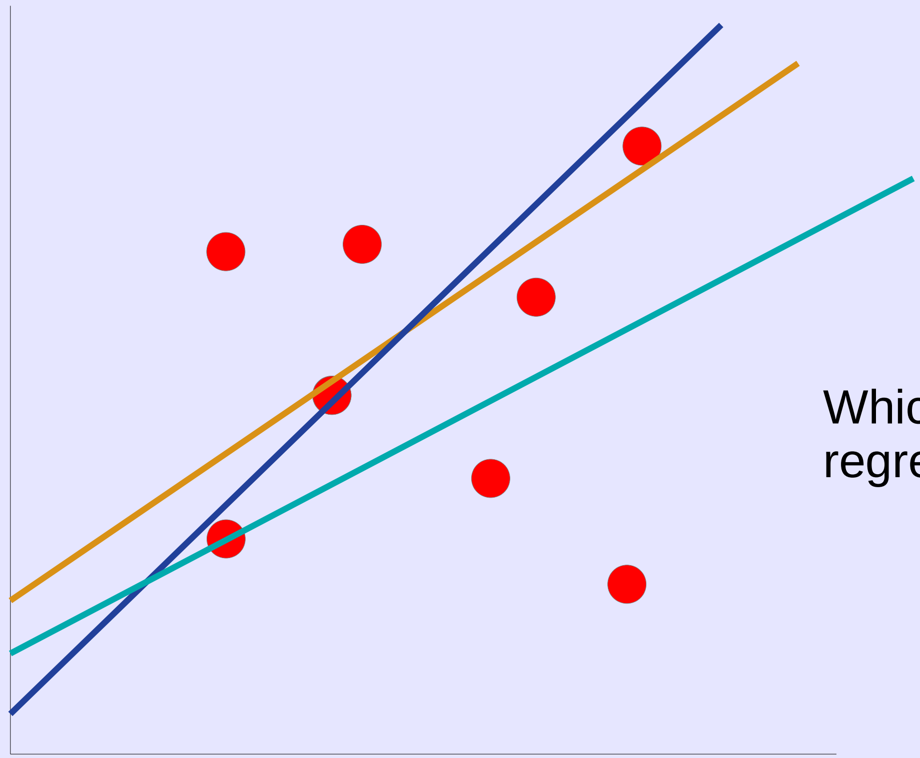
Linear regression: Theory

- Geometric interpretation



Linear regression: Theory

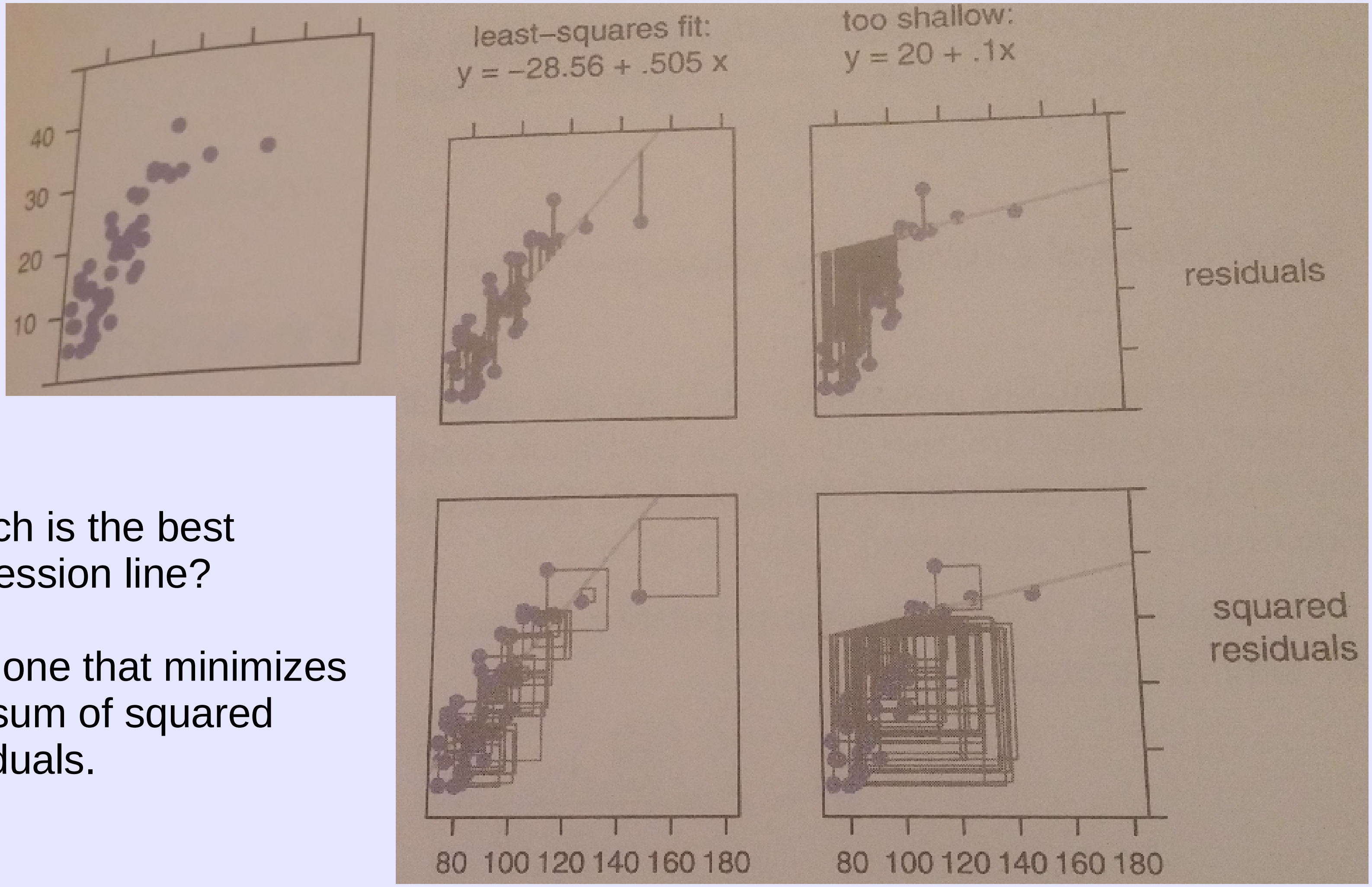
- Geometric interpretation



Which is the best regression line?

Linear regression: Theory

- Geometric interpretation



Which is the best regression line?

The one that minimizes the sum of squared residuals.

Linear regression: Theory

- So far, we have seen uni-variate regression with the following equation:

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

- Multi-variate regression can be generalized by the following equation:

  $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_n X_n + \varepsilon$

Linear regression: Example

- Empirical example: Advertising data consisting of 200x4 data frame.

	TV	radio	newspaper	sales
1	230.1	37.8	69.2	22.1
2	44.5	39.3	45.1	10.4
3	17.2	45.9	69.3	9.3
4	151.5	41.3	58.5	18.5
5	180.8	10.8	58.4	12.9
6	8.7	48.9	75.0	7.2
7	57.5	32.8	23.5	11.8
8	120.2	19.6	11.6	13.2
9	8.6	2.1	1.0	4.8
10	199.8	2.6	21.2	10.6

sales is the response variable (*Y*) (units are in thousands of some product)

TV, *radio* and *newspaper* are the predictors (*X*) (units in thousands of $)

Linear regression: Example

- Let's check the effect of radio advertising on the sales through linear regression:

```
> model.radio <- lm(sales ~ radio, data=df)
> summary(model.radio)

Call:
lm(formula = sales ~ radio, data = df)

Residuals:
     Min       1Q   Median       3Q      Max
-15.7305  -2.1324   0.7707   2.7775   8.1810

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  9.31164    0.56290  16.542   <2e-16 ***
radio        0.20250    0.02041   9.921   <2e-16 ***

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.275 on 198 degrees of freedom
Multiple R-squared:  0.332,	Adjusted R-squared:  0.3287
F-statistic: 98.42 on 1 and 198 DF,  p-value: < 2.2e-16
```

- Regression equation: sales = $\beta_0 + \beta_1$*radio
  = 9.312 + 0.203*radio