# Economic Analysis of Social Network (Part II)

Dr. Vinci Chow
vincichow@cuhk.edu.hk

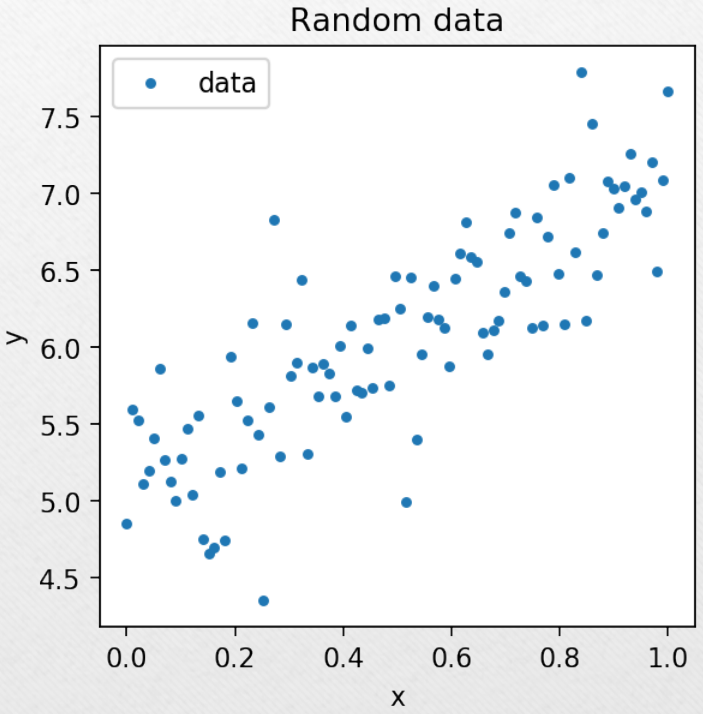
# What is Your Objective?

- Are you trying to:
  - Let the data tells you what relationship is contained within it?
    - **Unsupervised learning**
  - Trying to predict a continuous variable?
    - Supervised learning - **regression**
  - Trying to predict a discrete variable?
    - Supervised learning - **classification**

# Regression

- Ordinary Least Square (OLS) is the most common technique

$$\min_{\beta} \sum_{i} (y_i - \beta x_i)^2$$



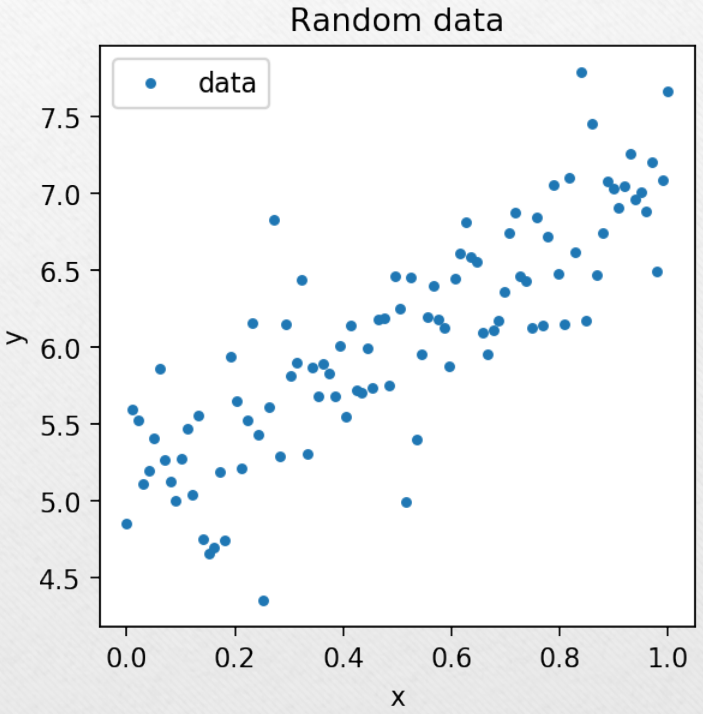
# Regression

- Ordinary Least Square (OLS) is the most common technique

$$\min_{\beta} \sum_{i} (y_i - \beta x_i)^2$$

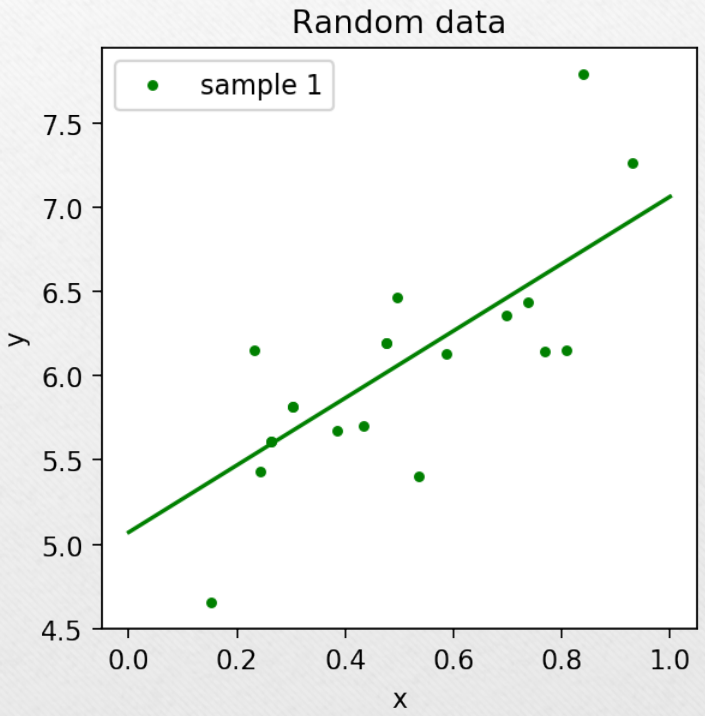
# Regression

- OLS is attractive when the objective is coefficient estimation because of its BLUE property
- It is, however, usually not ideal when it comes to making prediction



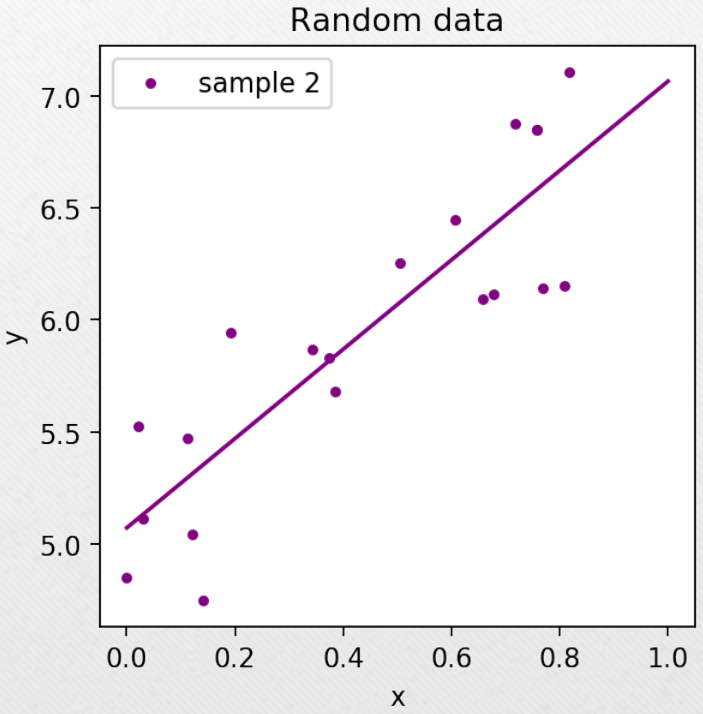
# Regression

- OLS is attractive when the objective is coefficient estimation because of its BLUE property
- It is, however, usually not ideal when it comes to making prediction



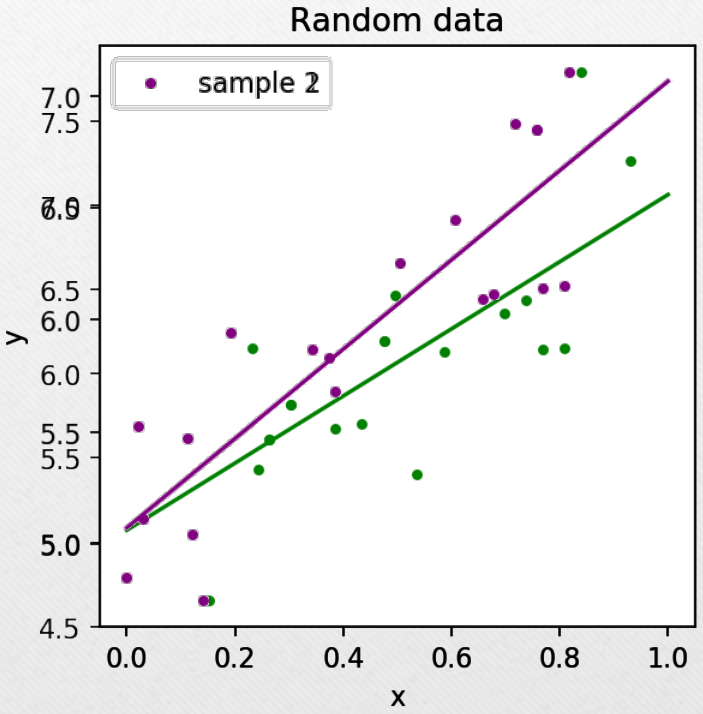
# Regression

- OLS is attractive when the objective is coefficient estimation because of its BLUE property
- It is, however, usually not ideal when it comes to making prediction



# Regression

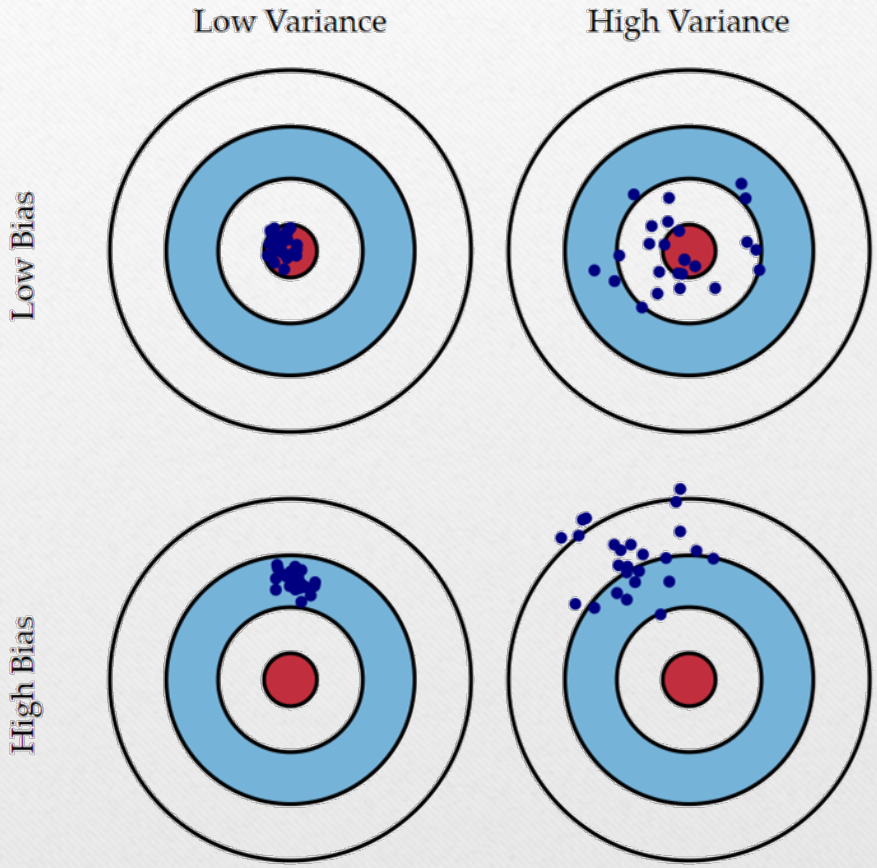
- OLS is attractive when the objective is coefficient estimation because of its BLUE property
- It is, however, usually not ideal when it comes to making prediction



# Unbiased Estimator

- The problem with OLS is that we are 100% focused on minimizing bias
- This makes our estimates highly dependent on the particular sample of data we have. In other words, we **overfit**
- How do we deal with this problem?

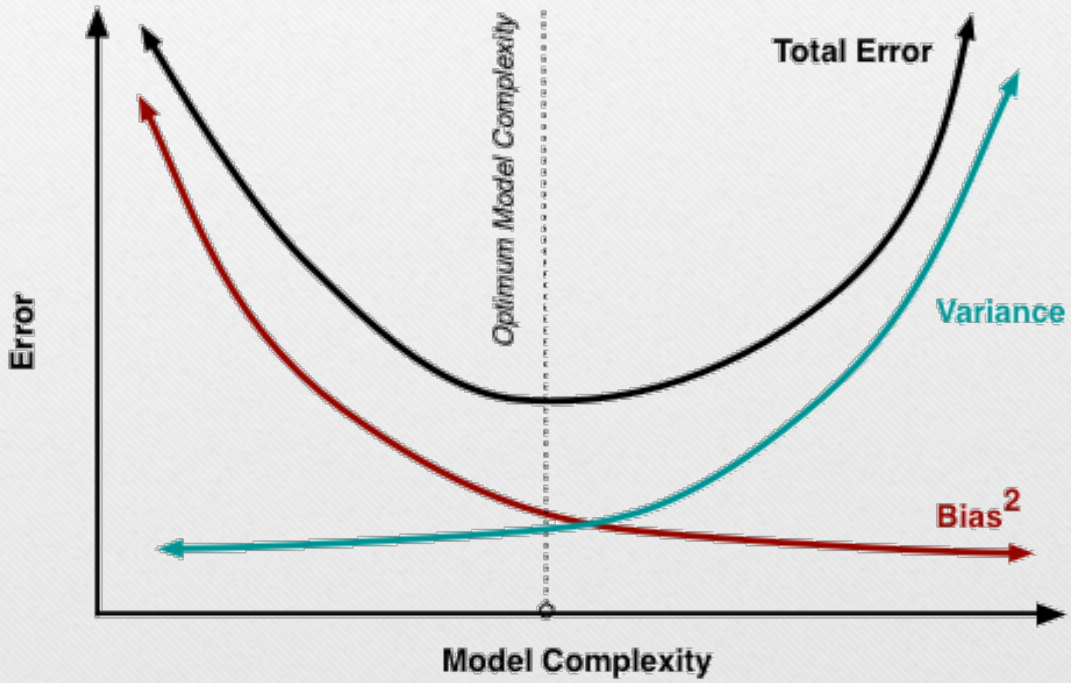
# Variance vs Bias




Source: http://scott.fortmann-roe.com/docs/BiasVariance.html

# Variance-Bias Tradeoff

- The tradeoff between a model's ability to minimize variance and to minimize bias




Source: http://scott.fortmann-roe.com/docs/BiasVariance.html

# Regularization

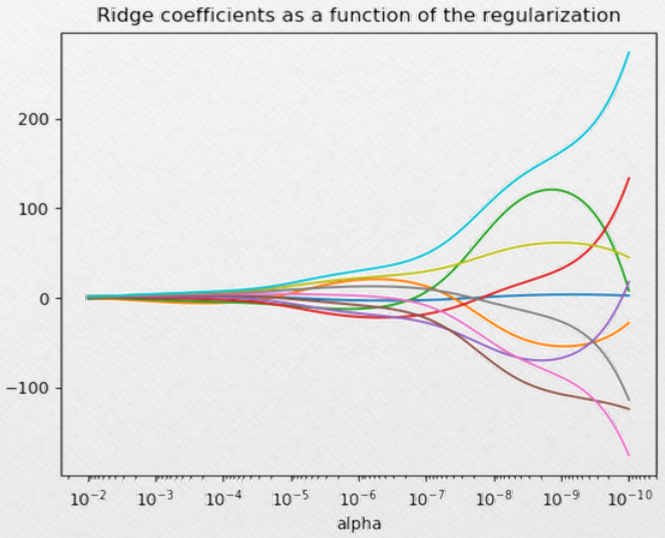
- **Regularization** penalizes large coefficients
- This reduce overfitting by making regression coefficients over different samples more similar to each other
  - In the most extreme case, all coefficients will be zero regardless of sample
- Regularized regression are biased, but they have smaller variance than OLS
- The two most regularized regressions are **ridge** and **lasso**

# Regularization

- Ridge Regression, also called **Least Square with L2-regularization**, have the following objective:

$$\min_{\beta}\left\{\sum_{i}(y_i-\beta x_i)^2+\alpha\sum_{k}\beta_k^2\right\}$$

- Ridge regression pushes **all** coefficients **towards zero**
- Stronger regularization (higher $\alpha$) leads to smaller coefficients



# Regularization

- Lasso Regression, also called **Least Square with L1-regularization**, have the following objective:

$$\min_{\beta} \left\{ \sum_{i} (y_i - \beta x_i)^2 + \alpha \sum_{k} |\beta_k| \right\}$$

- Lasso regression makes **some** coefficients **exactly zero**
- Stronger regularization (higher $\alpha$) leads to more coefficients becoming zero
- **Lasso can help you select variables**

# URL for Course Material

- econ1101.econ.cuhk.edu.hk:8000
- Resets every 8 hours. Let me know if you want a persistent version.