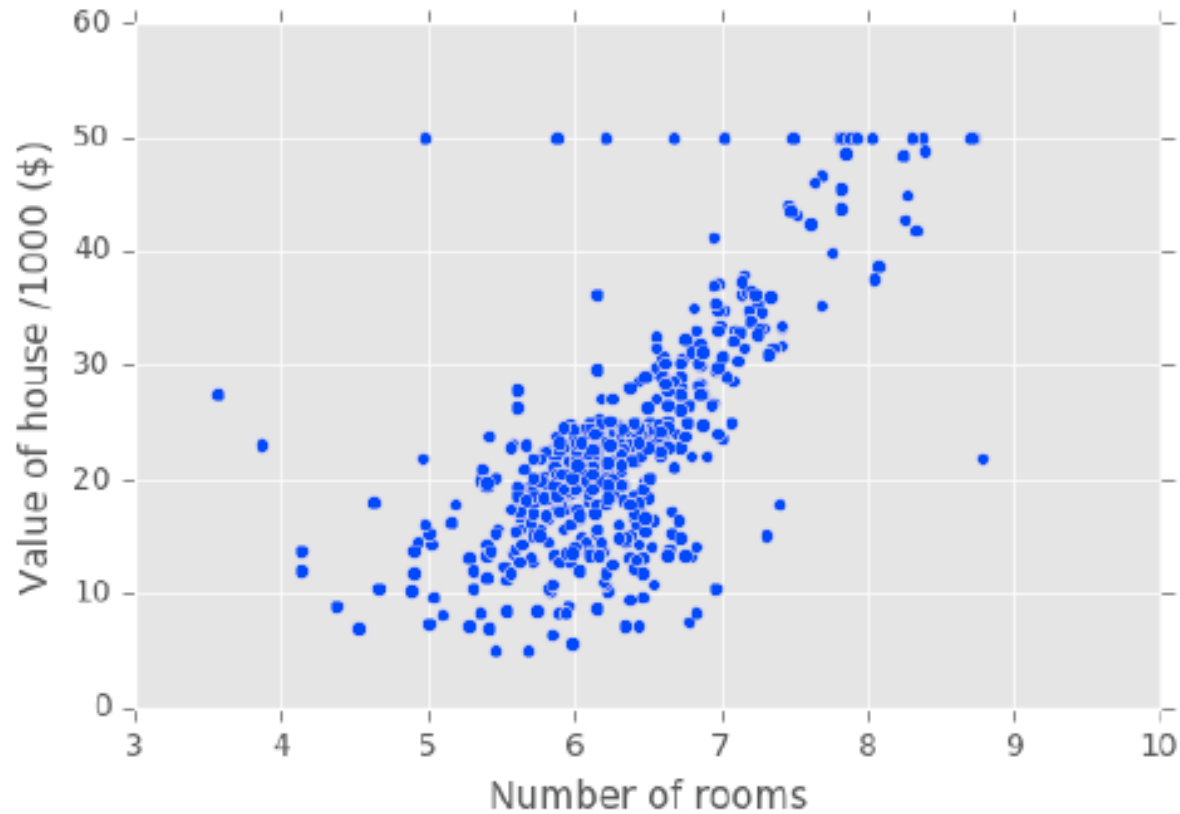
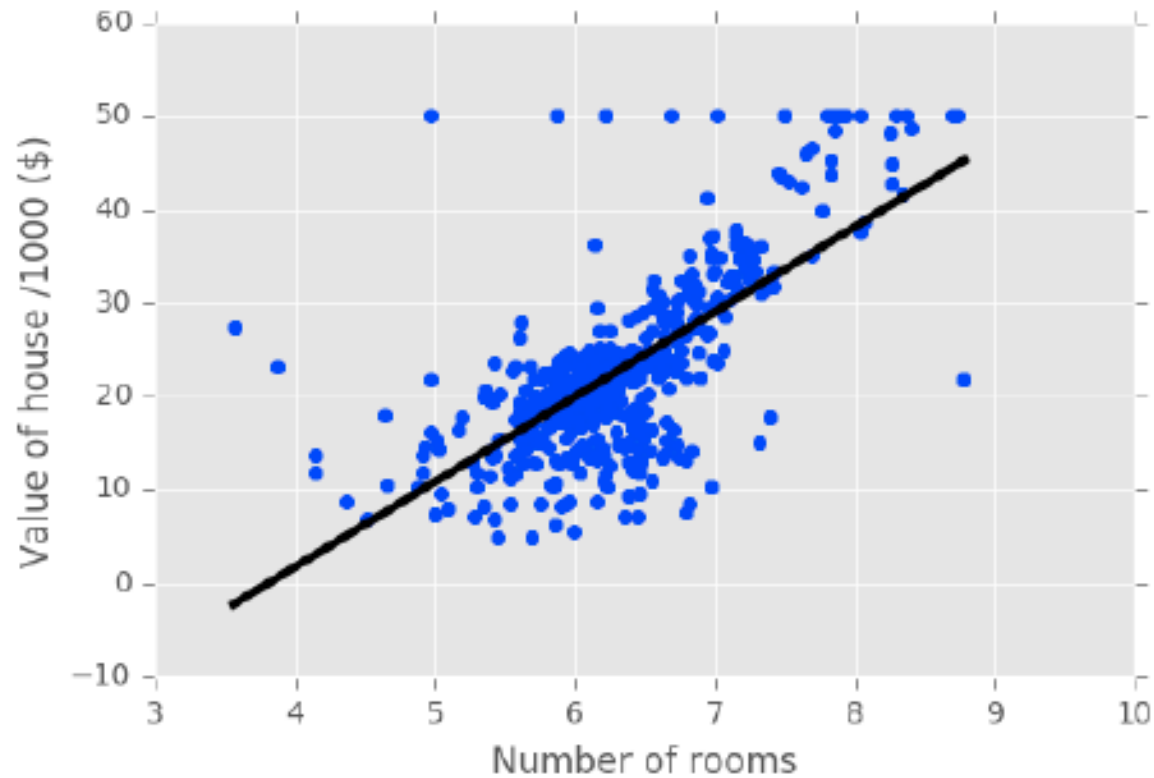


# Regression

# Plotting house value vs. number of rooms



# Fitting a regression model

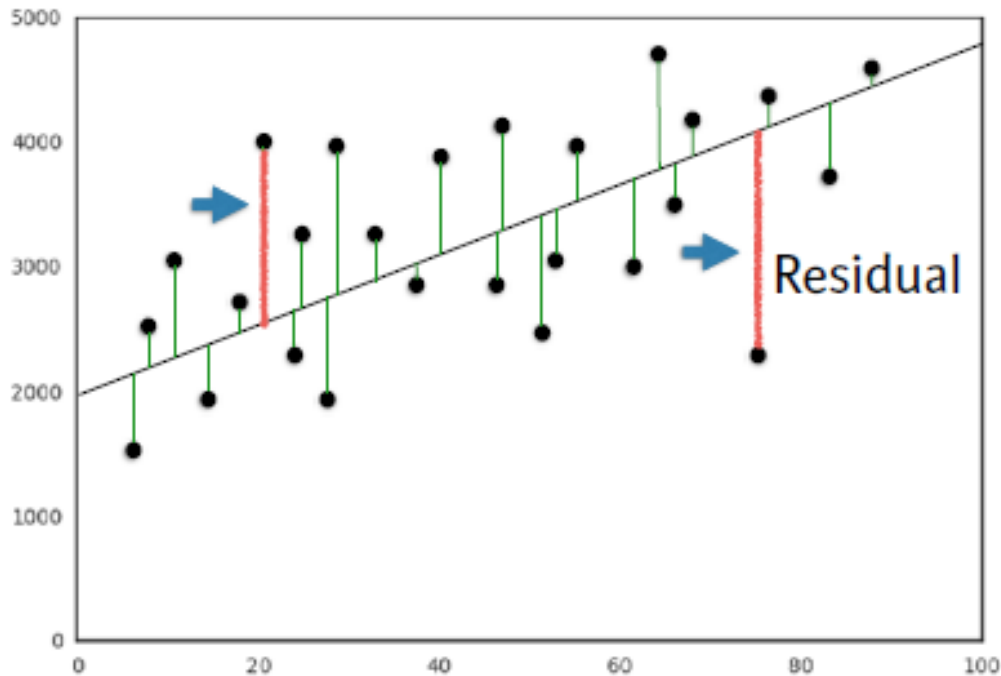


# Regression mechanics

- $y = ax + b$ 
  - $y$  = target
  - $x$  = single feature
  - $a, b$  = 모델의 매개 변수(parameters of model)
  - $a$ 와  $b$ 는 어떻게 선택합니까?
  - 주어진 라인에 대한 에러 함수 정의
  - 오류 기능을 최소화하는 선을 선택하십시오

# The loss function

- Ordinary least squares (OLS): Minimize sum of squares of residuals
  - 보통 최소 제곱 (OLS) : 잔차의 제곱의 합을 최소화합니다.



# Linear regression in higher dimensions

$$y = a_1x_1 + a_2x_2 + b$$

- 선형 회귀 모델을 여기에 맞추려면 다음을 수행함
  - 3 가지 변수를 지정해야함.

$$y = a_1x_1 + a_2x_2 + a_3x_3 + a_nx_n + b$$

- 각각 변수 a,b에 대한 계수를 지정해야함.
- Scikit-learn API는 정확히 같은 방식으로 작동함
  - 두 개의 배열 전달 : 기능 및 대상

# Cross-validation motivation

- 모델 성능은 데이터 분할 방식에 따라 다름.
- 모델의 일반화 능력을 대표하지 않음.
- 해결책 : 교차 검증!

# Cross-validation basics

Split 1	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Metric 1
Split 2	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Metric 2
Split 3	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Metric 3
Split 4	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Metric 4
Split 5	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Metric 5

Training data

Test data



# Cross-validation and model performance

- 5 folds = 5-fold CV
- 10 folds = 10-fold CV
- k folds = k-fold CV
- More folds = More computationally expensive

# Regularized regression

- 회귀 : 선형 회귀는 손실 함수를 최소화함.
- 각 특성 변수에 대한 계수를 선택함.
- 계수가 크면 오버 피터가 발생할 수 있음.
- 큰 계수 패널티 : 정규화

# Ridge regression

- Loss function = OLS loss function +  $\alpha * \sum_{i=1}^n a_i^2$
- Alpha: Parameter we need to choose
- Picking alpha here is similar to picking k in k-NN
- Hyperparameter tuning (More in Chapter 3)
- Alpha controls model complexity
- Alpha = 0: We get back OLS (Can lead to overfitting)
- Very high alpha: Can lead to underfitting

# Lasso regression

- Loss function = OLS loss function +  $\alpha * \sum_{i=1}^n |a_i|$

# Lasso regression for feature selection

- Can be used to select important features of a dataset
- Shrinks the coefficients of less important features to exactly 0

