Machine Learning

Module 3: Programming Homework

Ok-Ran Jeong

Fall, 2020

Dataset

- Dataset: insurance
  - 1338 records in insurance.csv
  - https://www.kaggle.com/mirichoi0218/insurance
  - Variables
    - You need to use a label encoder for categorical attributes

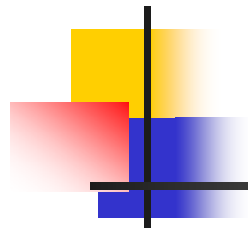| Variable	Description
Age	Age of primary beneficiary
Sex	Insurance contractor gender
BMI	Body mass index
Children	Number of children covered by health insurance
Smoker	Smoking (y/n)
Region	The beneficiary's residential area in the US
Charges	Individual medical costs billed by health insurance

Programming Homework 1

- Write a Python program for experimenting with Linear, Ridge and Lasso regularized regression models.

- For each model,
  - Do k-fold cross validation
    - k = 10
    - Scoring: 'neg_mean_squared_error'
  - Parameters
    - {'alpha': [1e-15, 1e-10, 1e-8, 1e-4, 1e-3, 1e-2, 1, 5, 10, 20]}

Programming Homework 2

- Write a Python program for experimenting with the Elastic Net regression model. (same dataset)

- For this model,

    - Do k-fold cross validation

        - k = 10

        - Scoring: 'neg_mean_squared_error'

    - Parameters

        - {'alpha': [1e-15, 1e-10, 1e-8, 1e-4, 1e-3, 1e-2, 1, 5, 10, 20]}

        - {'l1_ratio': [0.1, 0.3, 0.5, 0.7, 0.9]}

End of lab