Exercise no 3

FirstName LastName StudentNumber

Statistical Machine Learning

February 10, 2018

1 Optimization

Exercise 1.1 Find the solution of

- Orthogonal ridge: $S(\boldsymbol{\beta}) = \frac{1}{2n}\sum_{i=1}^{n}\sum_{j=1}^{p}(y_{ij} - \beta_j)^2 + \lambda\sum_{j=1}^{p}\beta_j^2$

  Hint: first solve $S_j = \frac{1}{2n}\sum_{i=1}^{n}(y_{ij} - \beta_j)^2 + \lambda\sum_{j=1}^{p}|\beta_j|$, and then generalize.

- Orthogonal lasso: $S(\boldsymbol{\beta}) = \frac{1}{2n}\sum_{i=1}^{n}\sum_{j=1}^{p}(y_{ij} - \beta_j)^2 + \lambda\sum_{j=1}^{p}|\beta_j|$

  Hint: first solve $S_j = \frac{1}{2n}\sum_{i=1}^{n}(y_{ij} - \beta_j)^2 + \lambda|\beta_j|$. Solve once for $\beta_j \geq 0$ and another time for $\beta_j < 0$. Put the pieces together. Then generalize.

- Ridge regression: minimize

$$S(\boldsymbol{\beta}) = \frac{1}{2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\top(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) + \lambda\boldsymbol{\beta}^\top\boldsymbol{\beta}$$

  Hint: use vector differentiation formulas of lecture01.

- Lasso: Write the coordinate descent algorithm for minimizing

$$S(\boldsymbol{\beta}) = \frac{1}{2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\top(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) + \lambda\sum_{j=1}^{p}|\beta_j|$$

Solution 1.1

Exercise 1.2 This extension of least squares allow you to model closed surfaces such as 3D scan of body or face.

Find the linearly constrained least squares estimator $(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\top(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$ subject to $\mathbf{T}\boldsymbol{\beta} = \mathbf{b}$ in which $\mathbf{T}$ and $\mathbf{b}$ both are known. Hint: use the Lagrangian dual.

How do you compute this estimator efficiently?

Solution 1.2

Exercise 1.3 Show that the eigen values of $\mathbf{A}+\lambda\mathbf{I}$ equals $\lambda_i+\lambda$ where λ_i's are the eigenvalues of $\mathbf{A}$. Use this result to argue that the ridge regression improves the condition number of $\mathbf{X}^\top\mathbf{X}$.

Solution 1.3

2 Mathematical Statistics

Exercise 2.1 Show the kernel density estimator $\hat{f}(y) = \frac{1}{n\lambda}\sum_{i=1}^{n} K(\frac{y_i - y}{\lambda})$ is a probability density, for any non-negative kernel that $\int_{-\infty}^{\infty} K(y)dy = 1$,

Hint : you must show $\hat{f}(y) \geq 0$ and $\int_{-\infty}^{\infty} \hat{f}(y)dy = 1$.

Exercise 2.2 A weighted linear model with weights $\mathbf{W}$ is called ordinary linear regression if $\mathbf{W} = \sigma^2\mathbf{I}$ for a known σ^2.

- Find the maximum likelihood estimator of $\boldsymbol{\beta}$ for the weighted linear regression. Weighted linear regression is $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ while $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \mathbf{W})$ and $\mathbf{W}_{n\times n}$ is the known variance covariance matrix of $\boldsymbol{\varepsilon}$.
- How do you compute the wighted least squares efficiently?
- What is $\mathbf{W}$ in $S(\boldsymbol{\beta}) = \sum_{i=1}^{n} a_i(y_i - \mathbf{x}_i^{\top}\boldsymbol{\beta})^2$?

3 Implementation

Exercise 3.1 How do you fit a weighted linear regression using a code that only fits the ordinary linear regression?

Hint: re-define $\mathbf{y}$ and $\mathbf{X}$ as a function of $\mathbf{W}$

Exercise 3.2 How do you fit ridge regression using a code that only fits the ordinary linear regression?

Hint: you must re-write the ridge estimator as a least squares problem and redefine a new (perhaps larger) $\mathbf{y}_\lambda$ and $\mathbf{X}_\lambda$ so that $(\mathbf{X}^\top\mathbf{X} + \lambda\mathbf{I})^{-1}\mathbf{X}^\top\mathbf{y} = (\mathbf{X}_\lambda^\top\mathbf{X}_\lambda)^{-1}\mathbf{X}_\lambda^\top\mathbf{y}_\lambda$.

Exercise 3.3 Chicago is well-known to be a crime city. Use R to discover where to buy a house in Chicago.

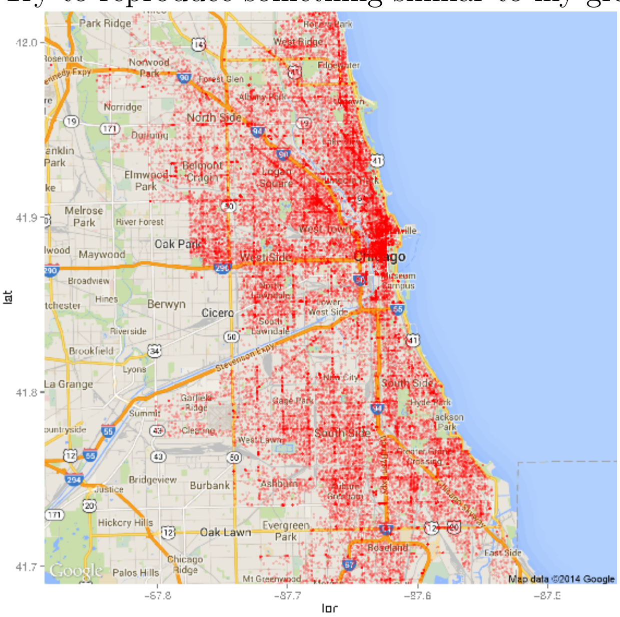
The crime data of Chicago are publicly available here

`https://data.cityofchicago.org/Public-Safety/Crimes-2001-to-present/ijzp-q8t2/data`

Take the subset of crimes committed in 2017 and plot a red dot (with enough transparency) for each *theft* appeared during 2017 on googlemap.

Hint: Use `get_map`, and `ggmap` functions from `ggmap` library to download the google map with appropriate zoom, longitude, and latitude. Then add red points using `geom_point` function of `ggplot2` library with transparency on your Chicago map. Using `ggplot` functions is not straightforward, check some simple examples first, before trying to overlay your points on the map.

Try to reproduce something similar to my graph (below).



Solution 3.1 Put your graph here.