Aprendizagem de Máquina I – Lista 02
Referente aos slides "02 Regressao Linear"

Prof. Hugo Carvalho

Fontes para os exercícios

- [ITSL] Gareth James, Daniela Witten, Trevor Hastie, Rob Tibshirani & Jonathan Taylor - *An Introduction to Statistical Learning, with Applications in Python* [baixe aqui]
- [AME] Rafael Izbicki & Tiago Mendonça dos Santos - *Aprendizado de Máquina: Uma Abordagem Estatística* [baixe aqui]

Questões do livro

- [ITSL] Capítulo 3
  - Questões conceituais: 3, 4
  - Questões aplicadas: 13
- [ITSL] Capítulo 6
  - Questões conceituais: 2, 3, 4
  - Questões aplicadas: 10 (*best subset selection* é você testar todas as 2^p combinações de atributos, analogamente ao que falamos em sala de aula para a regressão com penalização $\mathcal{P}_0$)

Questões avulsas

Questão 1: Encontre a solução analítica da regressão Ridge, minimizando em relação à variável $\boldsymbol{\beta}$ a função:

$$\|\mathbf{y} - \mathbb{X}\boldsymbol{\beta}\|^2 + \lambda\|\boldsymbol{\beta}\|_2^2.$$

A mesma estratégia que fizemos em sala para a regressão sem penalização funciona aqui. Para simplificar, assuma que o intercepto β_0 **não** está no vetor $\boldsymbol{\beta}$ – lembre-se que ele nunca é penalizado, por ser apenas um fator de deslocamento.

Questão 2: O objetivo dessa questão é encontrar a solução analítica da regressão Lasso no caso particular que a matriz de dados $\mathbb{X}$ é ortonormal, ou seja $\mathbb{X}^\intercal\mathbb{X} = \mathbb{I}$. Idealmente gostaríamos de minimizar, em relação à variável $\boldsymbol{\beta}$, a função:

$$\|\mathbf{y} - \mathbb{X}\boldsymbol{\beta}\|^2 + \lambda\|\boldsymbol{\beta}\|_1,$$

onde $\|\boldsymbol{\beta}\|_1 = \sum_{j=1}^{p} |\beta_j|$. Note que como a penalização não é uma função derivável, o procedimento usado na questão anterior não irá funcionar, de modo que outra alternativa deve ser buscada.

a) Consulte a Seção 2.4.1 da referência bibliográfica "Trevor Hastie, Robert Tibshirani & Martin Wainwright (2015) - *Statistical Learning with Sparsity: The Lasso and Generalizations*". Entenda o procedimento que lá é feito para encontrar a solução analítica do Lasso no caso univariado ($p = 1$).

b) Generalize o que você entendeu no item a) para o caso multivariado ($p > 1$). Aqui você precisará usar a hipótese de que $\mathbb{X}$ é uma matriz ortonormal.

Obs.: Se você está familiar com a noção de subgradiente, tente não utilizá-la! Treinar uma pequena gambiarra também é importante :-)