Regularization
正则化

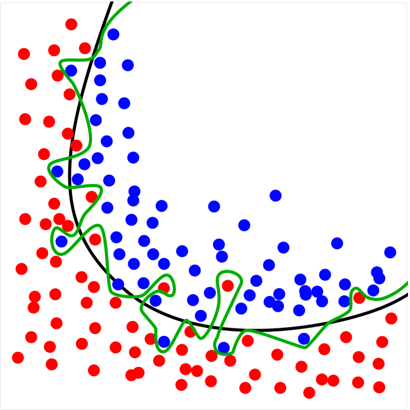
Sen LEI[1]

[1]UCSB
University of California, Santa Barbara
CA, U.S.

Aug, 2018

Summary

Background : Over-fitting



https://en.wikipedia.org/wiki/Overfitting

Some solutions

- 人为处理:
  - 模型选择算法
  - 人工选择保留哪些特征
- 正则化:
  - Lasso
  - Ridge
  - ...

正则化

OLS :

$$\hat{\beta} = \arg\min_{\beta} ||Y - X\beta||_2$$

正则化 :

$$\tilde{\beta} = \arg\min_{\beta} ||Y - X\beta||_2 + \lambda \, ||\beta||_{\alpha}$$

λ : 正则化参数

α : norm 参数

Definition

norm（范数）： 是一个具有“长度”概念的函数，其为向量空间内的所有向量赋予非零的正长度或大小

Definition

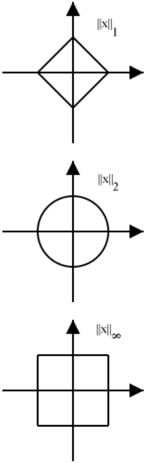
$\|\mathbf{x}\|_\alpha : \ \alpha = \ldots$

- $\|\mathbf{x}\|_1 = \sum_{i=1}^n |x_i|$ [Manhattan distance]
- $\|\mathbf{x}\|_2 = \sqrt{\sum_{i=1}^n x_i^2}$ [Euclidean distance]
- $\|\mathbf{x}\|_\infty = \max_i |x_i|$ [Chebyshev distance]
- $\|\mathbf{x}\|_{-\infty} = \min_i |x_i|$
- $\|\mathbf{x}\|_p = \left(\sum_{i=1}^n |x_i|^p\right)^{\frac{1}{p}}$

Unit circle

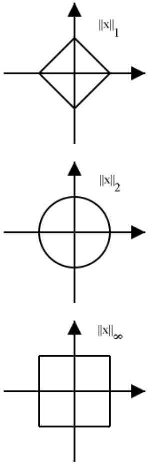
$\|\mathbf{x}\|_\alpha : \ \alpha = \ldots$

- $\|\mathbf{x}\|_1 = \sum_{i=1}^{n} |x_i|$
- $\|\mathbf{x}\|_2 = \sqrt{\sum_{i=1}^{n} x_i^2}$
- $\|\mathbf{x}\|_\infty = \max_i |x_i|$
- $\|\mathbf{x}\|_{-\infty} = \min_i |x_i|$
- $\|\mathbf{x}\|_p = \left(\sum_{i=1}^{n} |x_i|^p\right)^{\frac{1}{p}}$

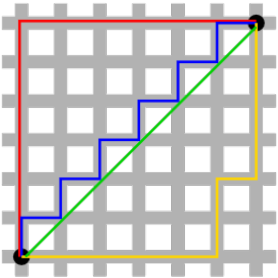


https://en.wikipedia.org/wiki/Norm_(mathematics)

L_1 norm & L_2 norm

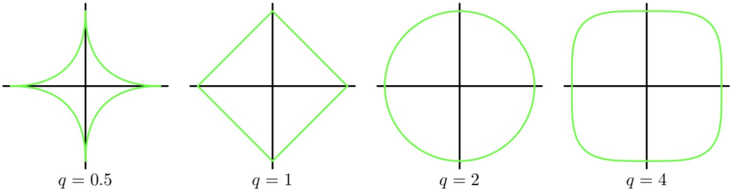


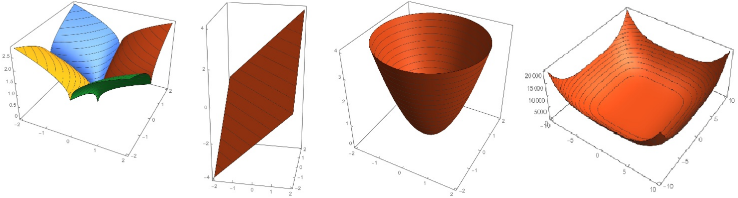
https://en.wikipedia.org/wiki/Norm_(mathematics)



https://en.wikipedia.org/wiki/Taxicab_geometry

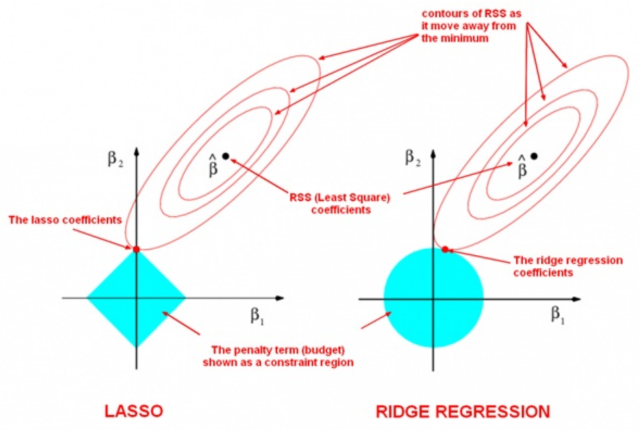
More





Pattern Recognition And Machine Learning - Springer 2006

Back to Lasso & Ridge Regression



https://stats.stackexchange.com/questions/347257/geometrical-interpretation-of-l1-regression

```
library(glmnet)

glmnet(
    x, y, family=c("gaussian","binomial","poisson",
                   "multinomial","cox","mgaussian"),
    weights, offset=NULL, alpha = 1, nlambda = 100,
    lambda.min.ratio = ifelse(nobs<nvars,0.01,0.0001),
    lambda=NULL, standardize = TRUE, intercept=TRUE,
    thresh = 1e-07, dfmax = nvars+1,
    pmax = min(dfmax*2+20, nvars), exclude,
    penalty.factor = rep(1, nvars),
    lower.limits=-Inf, upper.limits=Inf, maxit=100000,
    type.gaussian=ifelse(nvars<500,"covariance","naive"),
    type.logistic=c("Newton","modified.Newton"),
    standardize.response=FALSE,
    type.multinomial=c("ungrouped","grouped")
)
```

```
glmnet(x, y, family="gaussian", alpha = 1)
glmnet(x, y, family="gaussian", alpha = 0)
```

In R's package <glmnet>, the penalty is defined as :

$$(1-\alpha)/2\,\|\beta\|_2 + \alpha\,\|\beta\|_1$$

alpha=1 is the lasso penalty, and alpha=0 is the ridge penalty.

（注意：这里的 alpha 与 norm 定义中的下标 α 不同，两者没有直接联系）

```
import sklearn

sklearn.linear_model.ElasticNet(
    alpha=1.0, l1_ratio=0.5, fit_intercept=True, normalize=False,
    precompute=False, max_iter=1000, copy_X=True, tol=0.0001,
    warm_start=False, positive=False, random_state=None,
    selection='cyclic'
)
```

The ElasticNet mixing parameter, with 0 <= l1_ratio <= 1.
For l1_ratio = 0 the penalty is an L2 penalty.
For l1_ratio = 1 it is an L1 penalty.
For 0 < l1_ratio < 1, the penalty is a combination of L1 and L2.

(这里的 l1_ratio 类似 R <glmnet> 中的 alpha)

Find more on :
http://scikit-learn.org/stable/modules/generated/sklearn.linear_model.ElasticNet.html

Questions

问题？