Application of Linear Algebra in Machine Learning

Jason Wu / Olivia Yoon

Mentored by Edgar Jaramillo Rodriguez

University of California, Davis

June 16, 2021

Overview

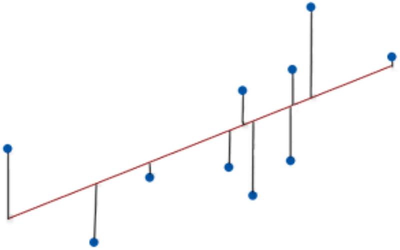
1. Regression Analysis
   - Linear Regression
   - Ridge Regression
   - Predictive Modeling in Python

2. Singular Value Decomposition
   - SVD Image Compression in Python

Ordinary Least-Squares (OLS)

Regression

Given the input vector X, we try to predict response variables and estimate relationships between the two.



Ordinary Least-Squares (OLS)

Problem

Given $x_1, \ldots, x_n \in \mathbb{R}^n$ and target variables $y_1, \ldots, y_n \in \mathbb{R}$ (n observations), we want the coefficients β for some $\beta \in \mathbb{R}^n$, such that minimizes

$$RSS(\beta) = \sum_{i=1}^{n} (y_i - x_i^T \beta)^2$$

Ordinary Least-Squares (OLS)

Geometric Model

$$\min_{\beta} \; = ||y - X\beta||_2^2$$

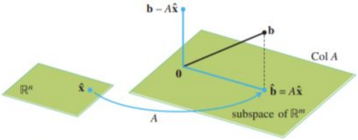


FIGURE 2 The least-squares solution $\hat{\mathbf{x}}$ is in $\mathbb{R}^n$.

Ordinary Least-Squares (OLS)

Solution

$\hat{\beta}$ is optimal if it satisfies

$$X^T X\beta = X^T y$$

if $X^T X$ is non-singular,

$$\hat{\beta} = (X^T X)^{-1} X^T y$$

Ridge Regression

Modification of Least Squares

Problem

Similar to OLS, Ridge Regression finds β that minimizes

$$||y - X\beta||^2 + \lambda||\beta||_2^2$$

Where $\lambda > 0$, a parameter we choose to control the amount of shrinkage.

Ridge Regression

Solution

Can be solved as OLS problems, where $\mathbb{I}$ is *nxn* identity matrix

$$(X^TX + \lambda\mathbb{I})\beta = X^Ty$$

Therefore,

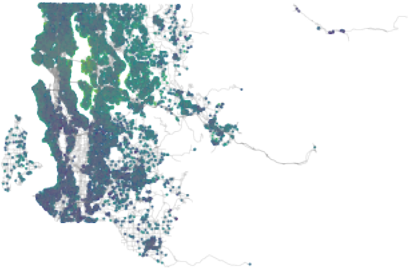
$$\hat{\beta} = (X^TX + \lambda\mathbb{I})^{-1}X^Ty$$

Predictive Modeling

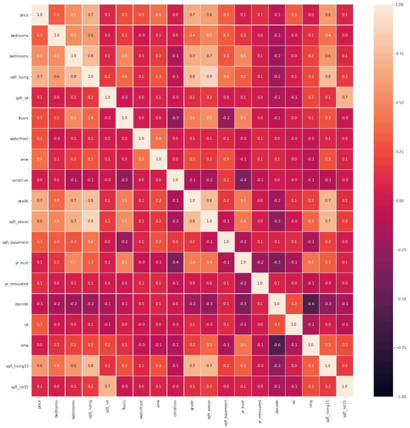
Implementation

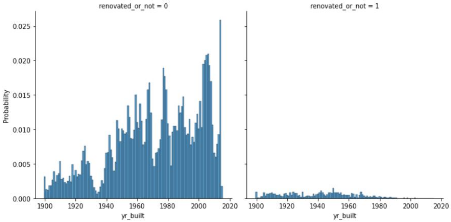
Using the data of 21,000 houses sold in King County, USA in 2014-2015, we create a model that can predict house prices from house features. After evaluating assumptions, such as multicollinearity and equal variance, and performing data-preprocessing, we test a model using linear regression and ridge regression from the sklearn package in Python.

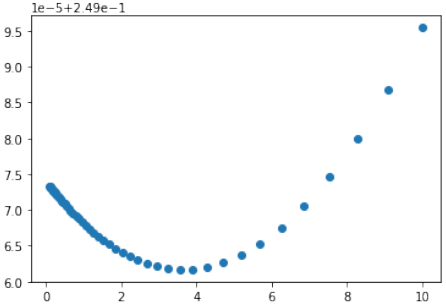
House Prices in King County



Predictive Modeling



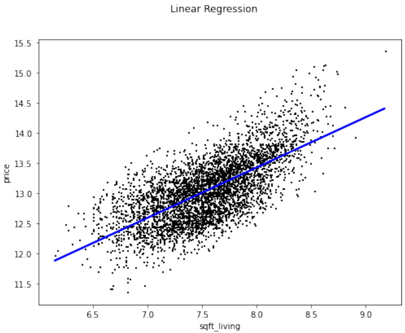


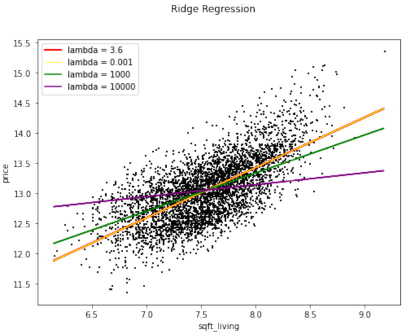


Predictive Modeling

Implementation

Ridge regression showed slightly higher accuracy (1-RMSE) in predicting the price.
Linear Regression: 0.750926
Ridge Regression: $\lambda = 3.6 : 0.750938$





Singular Value Decomposition

Problem

Given $A \in \mathbb{R}^{m \times n}$ of rank(n) and an integer $k << n$ find $A' \in \mathbb{R}^{m \times n}$ of rank(k) that minimizes

$$||A - A'||_F$$

Matrix Norm

Given an matrix A, the Frobenius Norm of A is

$$||A||_F = \sqrt{\sum_{i,j} A_{i,j}^2}$$

Singular Value Decomposition

Solution

Given a matrix A, its SVD is

$$A = U\Sigma V^T = \sum_{i=1}^{n} u_i \sigma_i v_i^T$$

where U and V are orthonormal and Σ is a rectangle diagonal matrix. The best rank(k) approximation is

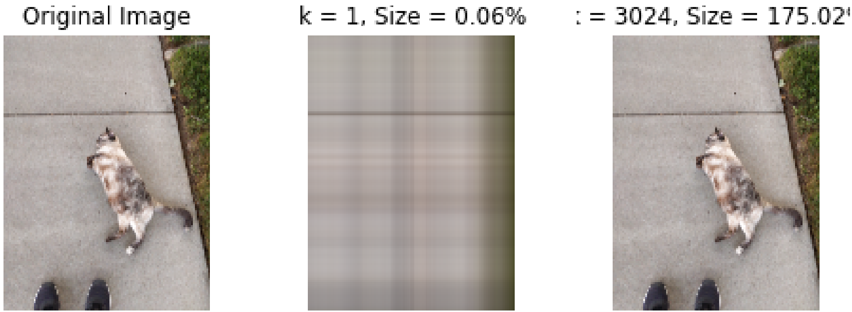
$$A' = \sum_{i=1}^{k} u_i \sigma_i v_i^T$$

where σ_i are the singular values of A in nonincreasing order.

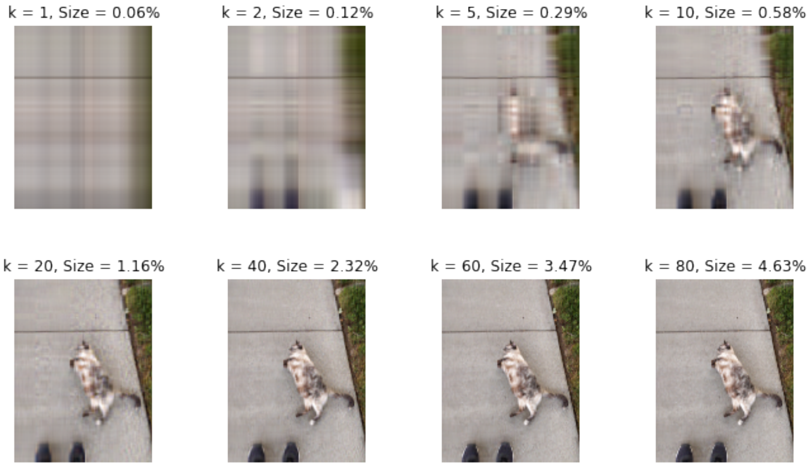
SVD Image Compression

Application

Images can be treated as matrices of pixel values. With ever higher resolution cameras, these matrices can be quite large and take up a lot of storage and bandwidth. The use of SVD allows for efficient compression of the images.



SVD Image Compression



References

Jim Frost
https://statisticsbyjim.com/glossary/ordinary-least-squares/

Frank Cleary
https://www.frankcleary.com/svdimage/

David C. Lay
Linear Algebra and its Applications

Trevor Hastie, Rober Tibshirani, Jerome H. Friedman
The Elements of Statistical Learning

Thank You! Any Questions?