Regularization

—

Najiyah Khan
DSI
January 11, 2017

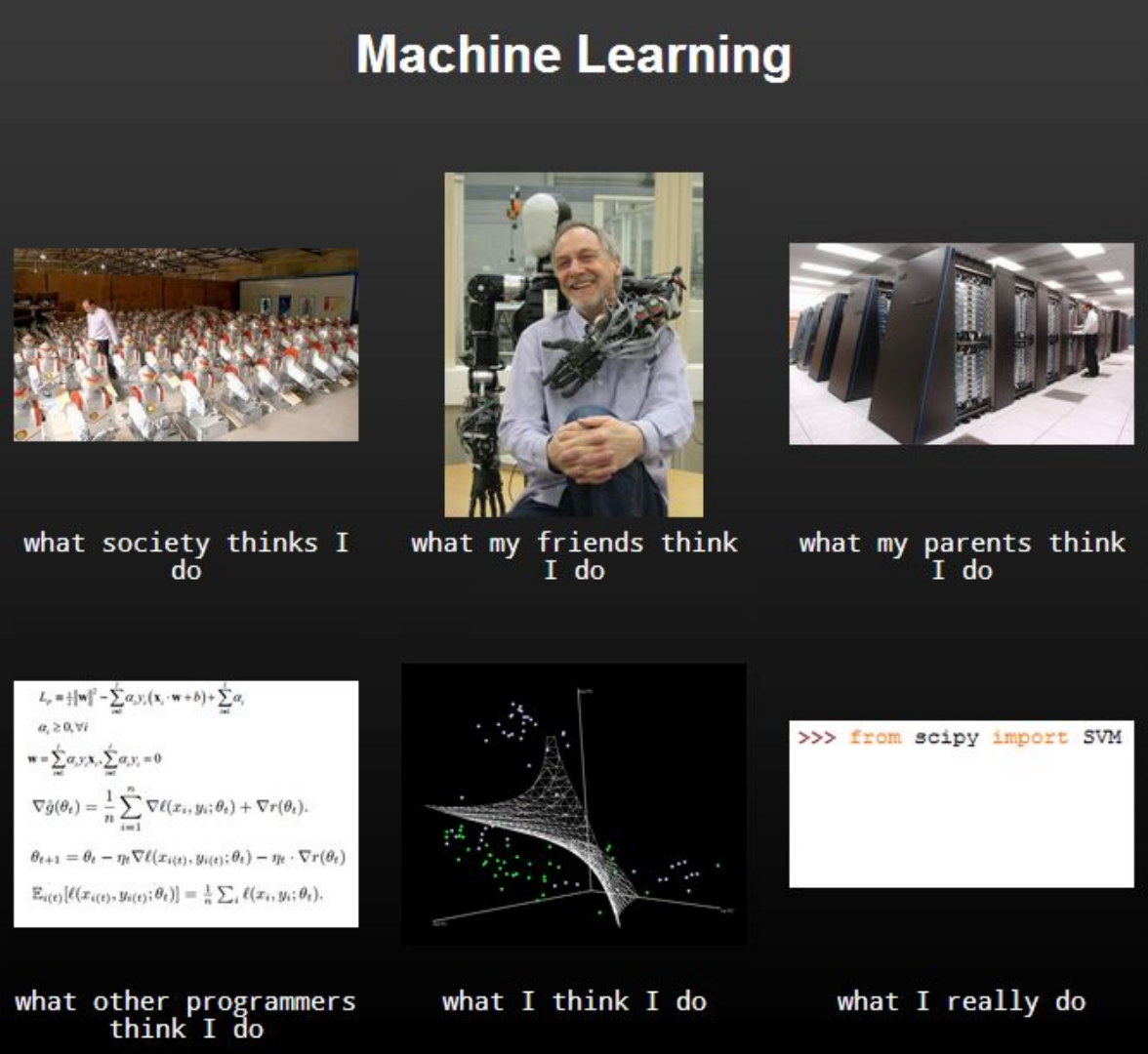
Agenda

- Machine Learning
- Regularization
- Lasso
- Ridge
- Examples
- Resources
- Conclusion
- Questions

Machine Learning

"..gives computers the ability to learn without being explicitly programmed" (Arthur Samuel, 1959).

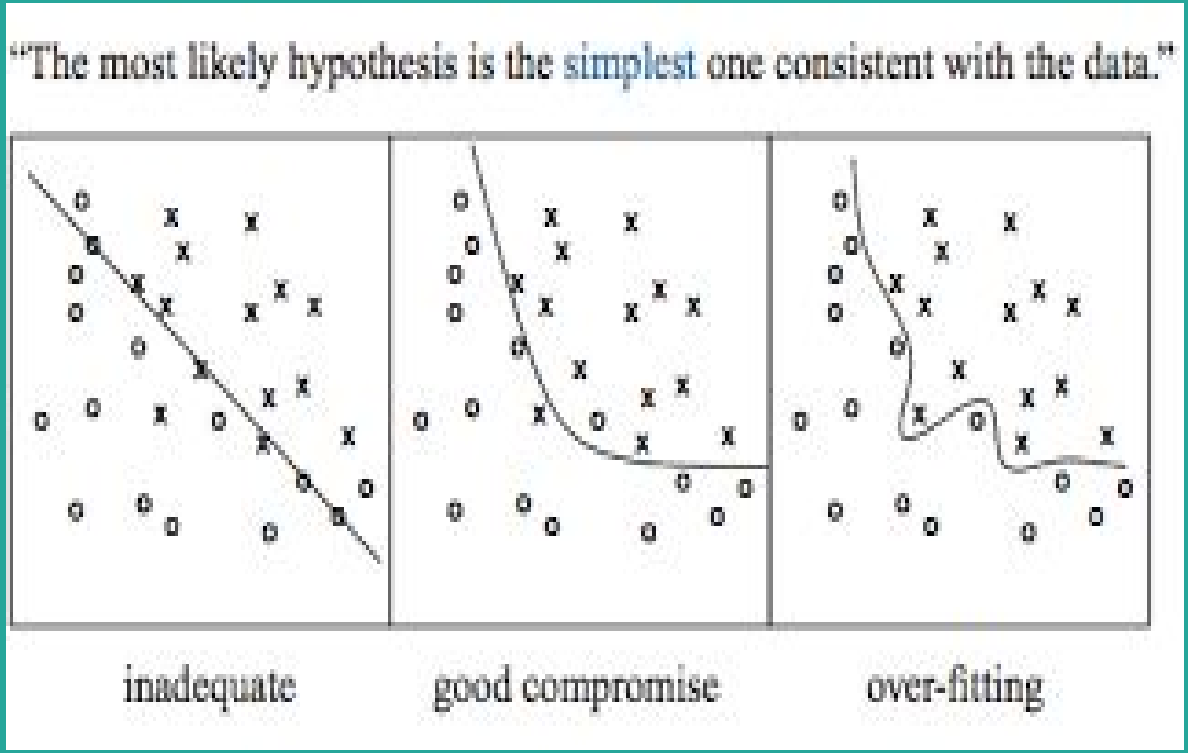
- Self-driving Google Car.
- Filtering for email spam.
- Netflix movie suggestions.



Definition of Regularization

- It is a technique that counters the problem of overfitting by introducing tuning parameters that ease the full power of the features/variables in your dataset.
- Overfitting occurs when your predictive model aligns itself too closely with your data and not the real world.
- Two of the most used tuning parameters are:
  - L1 - Lasso
  - L2 - Ridge
- Other parameters:
  - Least Angle Regression (LARS) -
  - Elastic Net - regularized regression methods that linearly combines the penalties of the Lasso and Ridge tuning parameters. (https://en.wikipedia.org/wiki/Elastic_net_regularization)

Overfitting



Lasso Regression (L1)

- This method shrinks the value of the coefficients (toward zero) using the sum of the absolute value of the coefficients.
- Lasso is preferred if we believe many features are irrelevant or we have a sparse model.

Ridge Regression (L2)

- We are again imposing a penalty on the coefficient size.
- We are shrinking the squares of the coefficients instead of using the sum of the absolute value-which was done in Lasso (L1).

Examples

Predicting number of bikes rented every hour.
DC Dept. of Transportation

- FeelsLikeTemp
- Humidity
- Windspeed
- Non-registered renters
- Registered Renters
- Number of total rentals
- DateTime
- Season
- Holiday
- WorkingDay
- Weather

Predicting Salary

- Age
- Sex
- Location
- Experience
- Race
- Family Status
- Education

Resources

- Analytics Vidhya
  - https://www.analyticsvidhya.com/blog/2016/01/complete-tutorial-ridge-lasso-regression-python/
- Stats website
  - http://www.statisticshowto.com/regularization/
- Machine Learning through Coursera/Stanford taught by Andrew Ng
  - https://www.coursera.org/learn/machine-learning

Conclusion and Questions

"The goal is to turn data into information, and information into insight." – Carly Fiorina, former chief executive of Hewlett-Packard Company