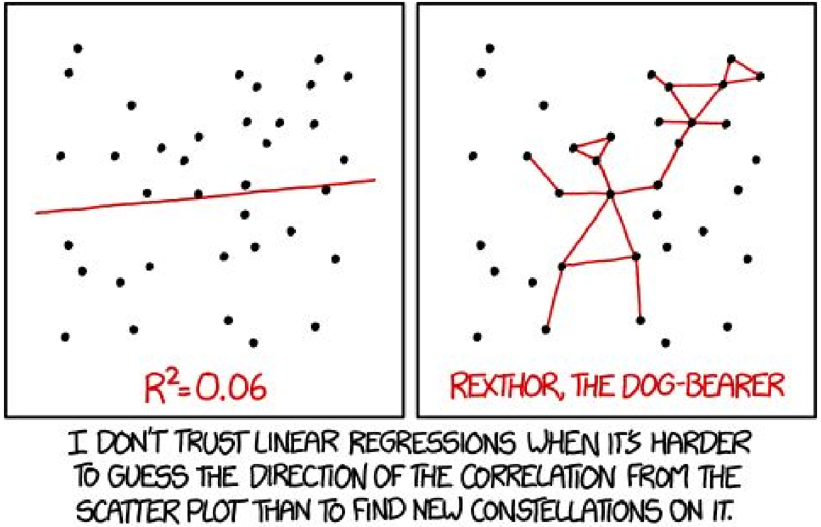
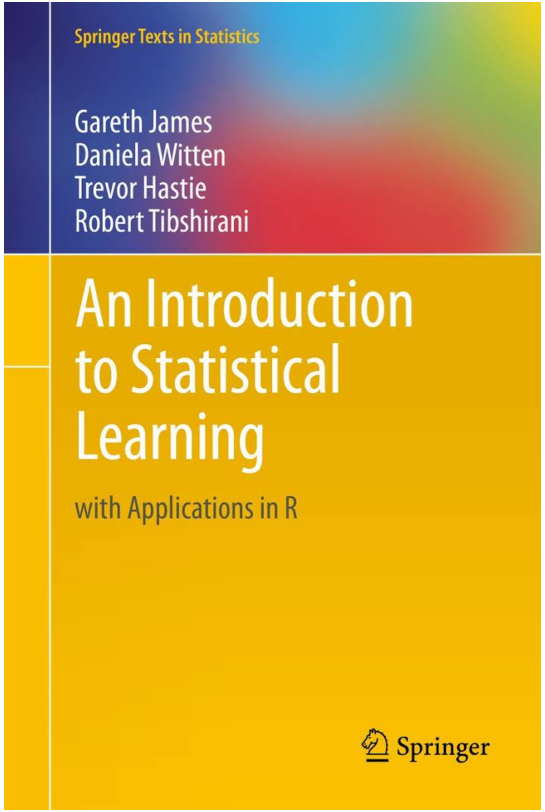
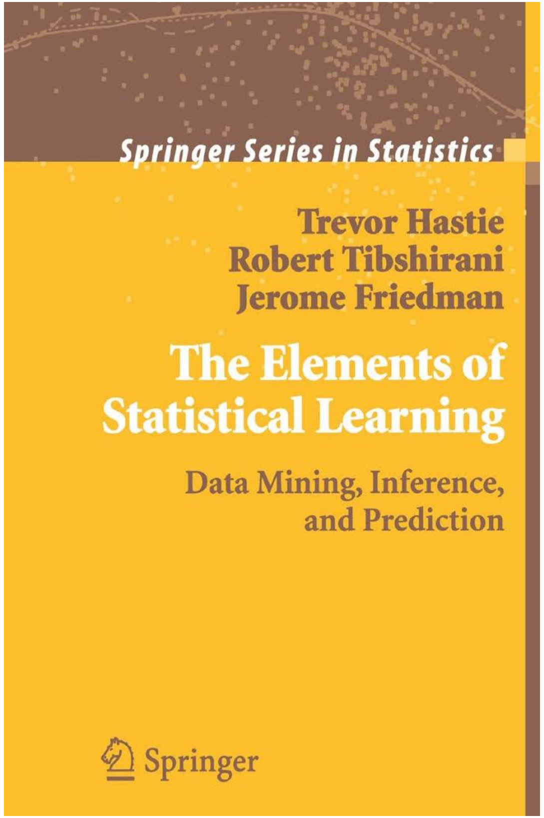
Linear Regression (2)



Seoul AI Meetup

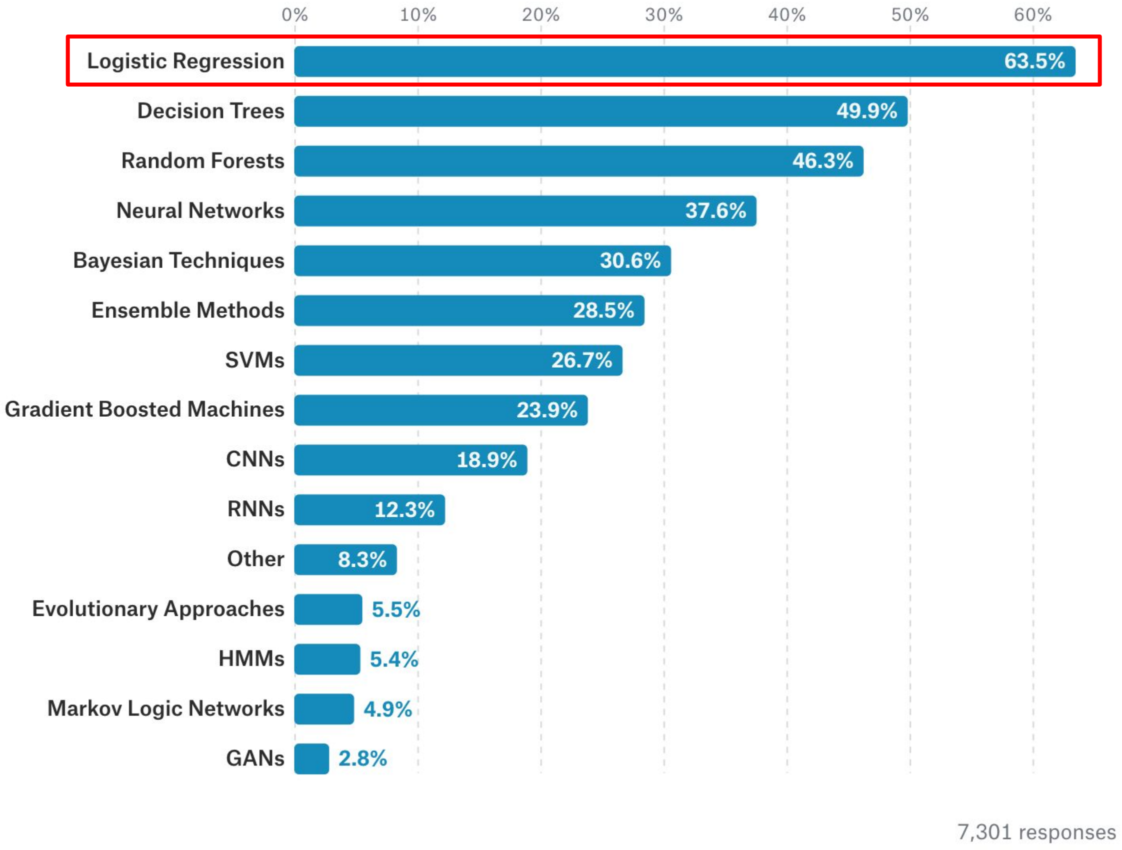
Martin Kersner, 2017/11/11





http://www-bcf.usc.edu/~gareth/ISL/

https://web.stanford.edu/~hastie/ElemStatLearn/



https://www.kaggle.com/surveys/2017

Linear Regression Methods

Introduction

- Ordinary Least Squares Regression

Basic methods

- **Polynomial Regression**
- **Locally Weighted Linear Regression**

Shrinkage methods

- Ridge Regression
- Lasso
- Forward Stagewise Regression

Others

- TensorFlow Lattice

Ordinary Least Squares

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

$$m > n$$

http://seoulai.com/presentations/Linear_Regression_1.pdf

Locally Weighted Linear Regression

Locally Weighted Linear Regression

Linear regression usually underfits data.

Not all data points are relevant for prediction.

Give more weight (W) to data points near data point of interest.

Drawback: Uses full dataset each time a prediction is needed!

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^\top \boldsymbol{W} \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{W} \boldsymbol{y}$$

Locally Weighted Linear Regression

Gaussian kernel

Closer data points to data point of interest will have larger influence on prediction than farther ones.

Distance between data points measured by Gaussian function*.

$$w_{i,i} = \exp \frac{(x_i - \boxed{x})^2}{-2k^2}$$

Point of interest

***k* hyperparameter**

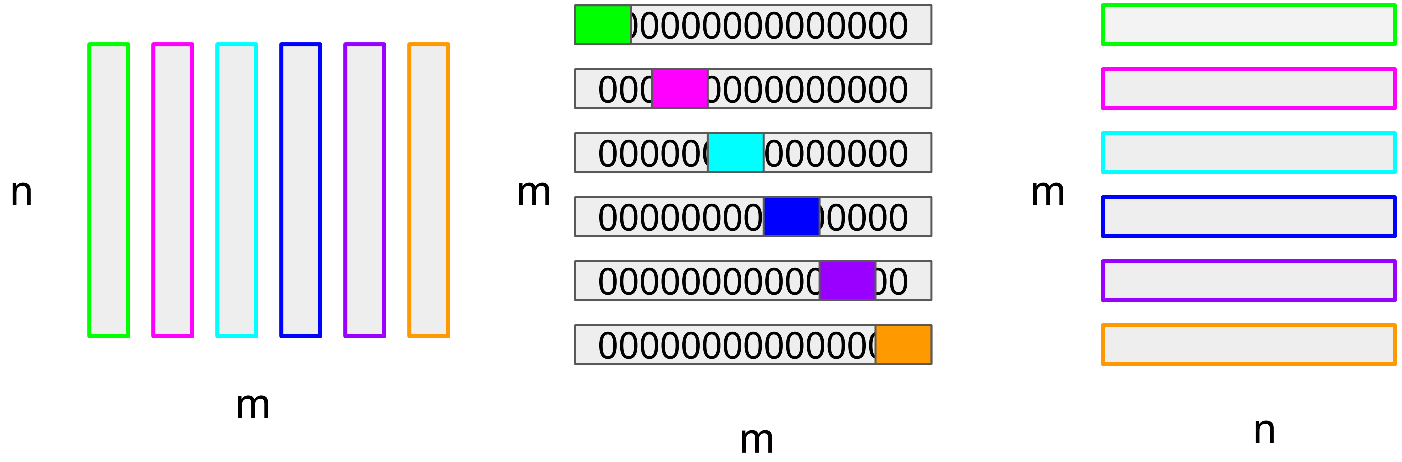
Defines how much to weigh nearby points.

When ***k*** is higher further data points will have influence on prediction.

* https://en.wikipedia.org/wiki/Gaussian_function

Locally Weighted Linear Regression

$$\boldsymbol{X}^\top \boldsymbol{W} \boldsymbol{X}$$

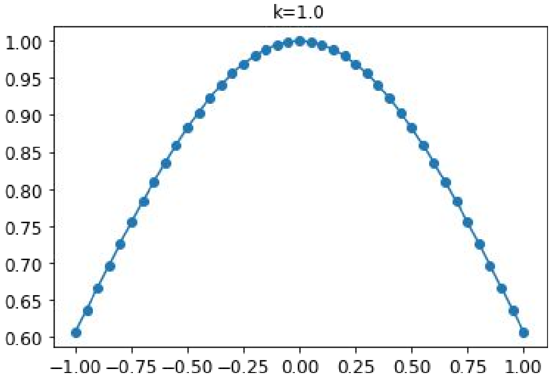


Computation of $\boldsymbol{X}^\top \boldsymbol{W} \boldsymbol{y}$ is based on the same principle.

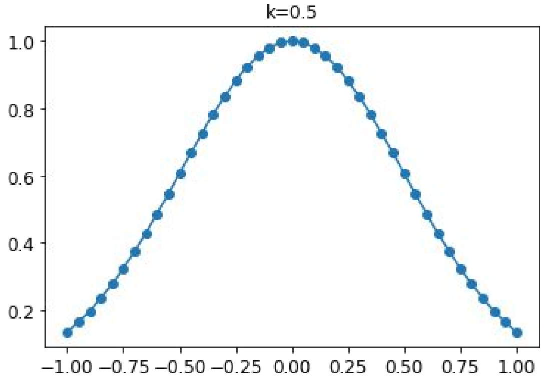
Locally Weighted Linear Regression

Gaussian kernel

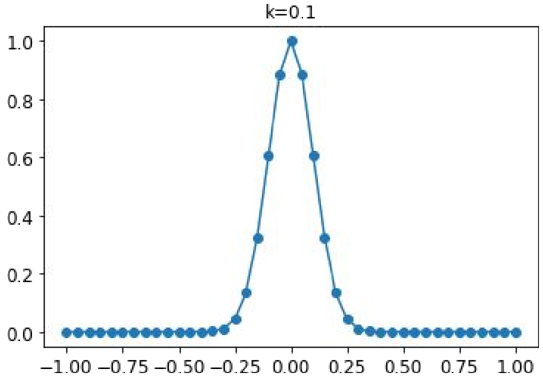
k=1.0



k=0.5



k=0.1



Polynomial Regression

Polynomial

Expression that can be built from constants and indeterminates by means of addition, multiplication and exponentiation to non-negative integer power.*

$$\boxed{a_0} + \boxed{a_1}x + \boxed{a_2}\boxed{x}^{\boxed{2}} + \cdots + \underbrace{\boxed{\boxed{a_{n-1}}\boxed{x}^{\boxed{n-1}}}}_{\text{Term}} + \boxed{a_n}\boxed{x}^{\boxed{n}}$$

Degree of indeterminate = The exponent of indeterminate.

Degree of term = The sum of degrees of indeterminates.

Degree of polynomial = The highest degree of terms.

Constant , Linear, Quadratic, Cubic, ...

* https://en.wikipedia.org/wiki/Polynomial#Definition

Polynomial Regression

Special case of Multiple Linear Regression.

Relationship between the independent variable X and the dependent variable y is modelled as an ***n*th degree polynomial** in X.*

Enables us to fit a nonlinear model to the data, but as a statistical estimation problem it is linear.

$$\boldsymbol{X} \in R^{m,n} \rightarrow \phi(\boldsymbol{X}) \in R^{m,d_\phi}$$

$$\hat{\boldsymbol{\beta}} = (\phi(\mathbf{X})^\mathsf{T}\phi(\mathbf{X}))^{-1}\phi(\mathbf{X})^\mathsf{T}\boldsymbol{y}$$

* https://en.wikipedia.org/wiki/Polynomial_regression

Polynomial Regression

Example

Add terms with degrees greater than one to the model.

$$a \xrightarrow{\phi_2} a, a^2$$

$$a \xrightarrow{\phi_3} a, a^2, a^3$$

$$a, b \xrightarrow{\phi_2} a, b, a^2, b^2, ab$$

$$a, b \xrightarrow{\phi_3} a, b, a^2, b^2, ab, a^2b, ab^2, a^3, b^3$$

It is unusual to use degree greater than 3 or 4.

Polynomial Regression

Independent variable are not assumed to be random, but fixed, therefore we can transform them in arbitrary ways.

Basis functions

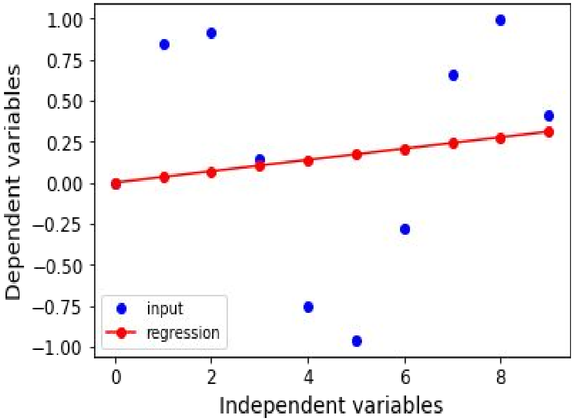
A fixed set of nonlinear functions that are used to transform the value(s) of a data point(s).

Other basis functions such as splines, radial basis functions, and wavelets can be utilized.

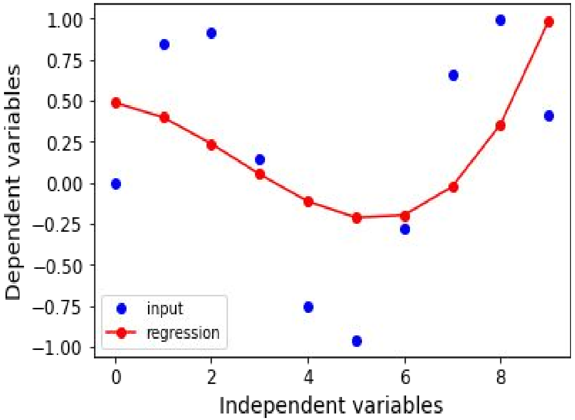
Polynomial Regression

Underfitting & Overfitting

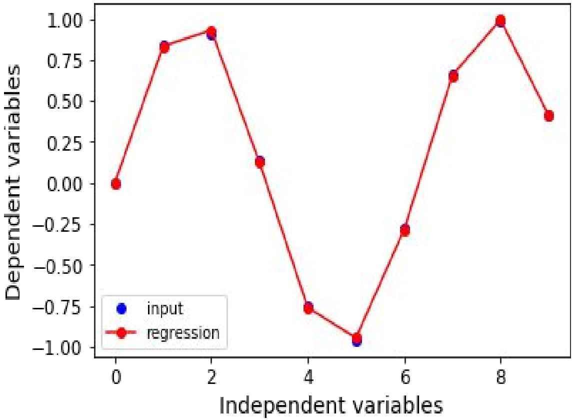
OLS



Degree = 3



Degree = 6



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

Quantitative vs Qualitative Features

Quantitative Features

Two levels

$$x_i = \begin{cases} 1 & \text{if condition is True} \\ 0 & \text{if condition is False} \end{cases}$$

name	female
Martin	0
Yoonju	1

Three levels

$$x_1 i = \begin{cases} 1 & \text{if condition1 is True} \\ 0 & \text{if condition1 is False} \end{cases}$$

$$x_2 i = \begin{cases} 1 & \text{if condition2 is True} \\ 0 & \text{if condition2 is False} \end{cases}$$

name	American	Korean
Raj	1	0
Yoonju	0	1
Martin	0	0

Summary

Models are all linear from the point of view of estimation.

Fit a nonlinear relationship.

Both can be fit using Ordinary Least Squares.

References

Polynomial Regression

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

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