Statistical Learning Survey

Abraham Neuwirth

May 29, 2017

Contents

1 Introduction

Christopher Bishop (Bishop 2006) in his seminal book "Pattern Recognition and Machine Learning", introduces the field of statistical learning with a classical story:

> [T]he extensive astronomical observations of Tycho Brahe in the 16th century allowed Johannes Kepler to discover the empirical laws of planetary motion, which in turn provided a springboard for the development of classical mechanics.

It's the archtype of statistical learning success stories. Lot's of data, brilliant minds, and a model to illuminate and explain it all.

2 Data Representation

2.1 Vectors

2.2 Matrices

2.3 Tensors

3 Supervised Learning

Lemma 3.1. *This is a lemma*

3.1 Linear Regression

Of the most simplest and common statistical learning methods, linear regression is the most common. It is also one of the oldest, having been discovered (Kopf 2015) independently by Adrien-Marie Legendre in 1805 (Legendre 1805), and then again by Carl Friedrich Gauss in 1809 (**???**).

In the most simplest terms, the objective of a regression is to find the best line to approxiamte a given data set consisting of many points (x_i, y_i), where x_i is the independent variable and y_i is the dependednt variable.

The objective function is to find estimated values for x, such that the distance between the actual y and $\hat{y}$ is mimimized. More formally:
Definition 3.1. A residual r_i is the difference between the actual value of the dependent variable and the the value predicted by the model such that $r_i = y_i - f(x_i, \beta)$

The least squares method optimizes the sum of the squares of the residuals:

$$S = \sum_{i=1}^{n} r_i^2$$

As an example, we'll use the data from (**???**) to illust

Now, the regression draws a straight line through the points 3.2 ok

The form of regression published by Legendre and Gauss was what is now called Least Squares regression.

3.2 Support Vector Machines

4 Unsupervised Learning

4.1 Classification

4.2 Naive Bayes

4.3 K-Nearest Neighbors

5 Regularlization

Overfitting is a problem. regularlization penalizes biggest predictors. "Regularization can be accomplished by restricting the hypothesis space $\mathcal{H}$"

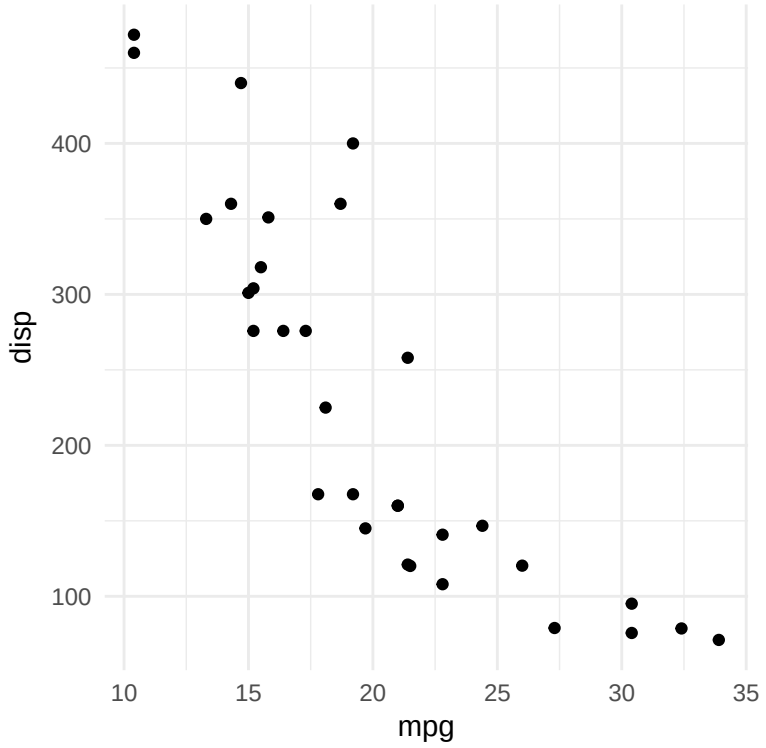

Figure 3.1: the data

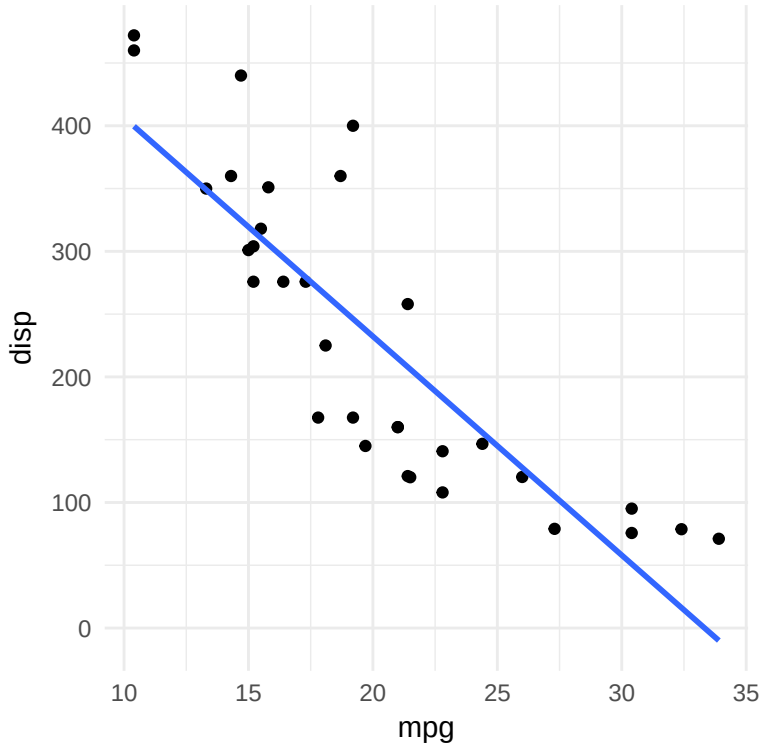

Figure 3.2: regression

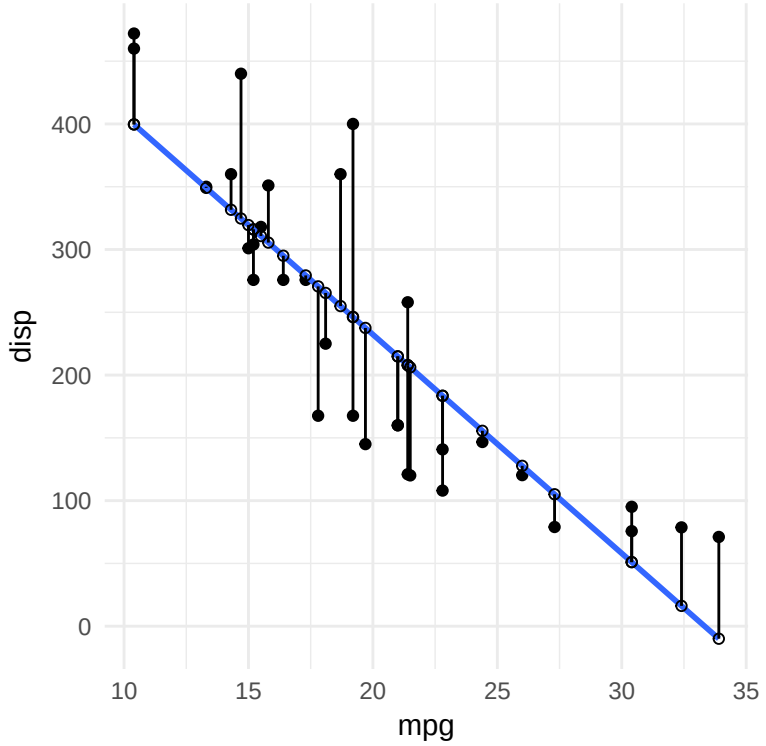

Figure 3.3: residuals

5.1 Tikhonov regularization (Ridge Regression)

5.2 Lasso Regression

5.3 Principal Components

Dimensionality reduciton

6 Reinforcment Learning

7 Deep Learning

This is a level 1 heading

7.1 Level 2

Blah blah (se also James et al. 2014 ch. 1) and Hastie, Tibshirani, and Friedman (2001).

7.1.1 equations

This s an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see http://rmarkdown.rstudio.com. See equation (7.1) there's also (7.2)

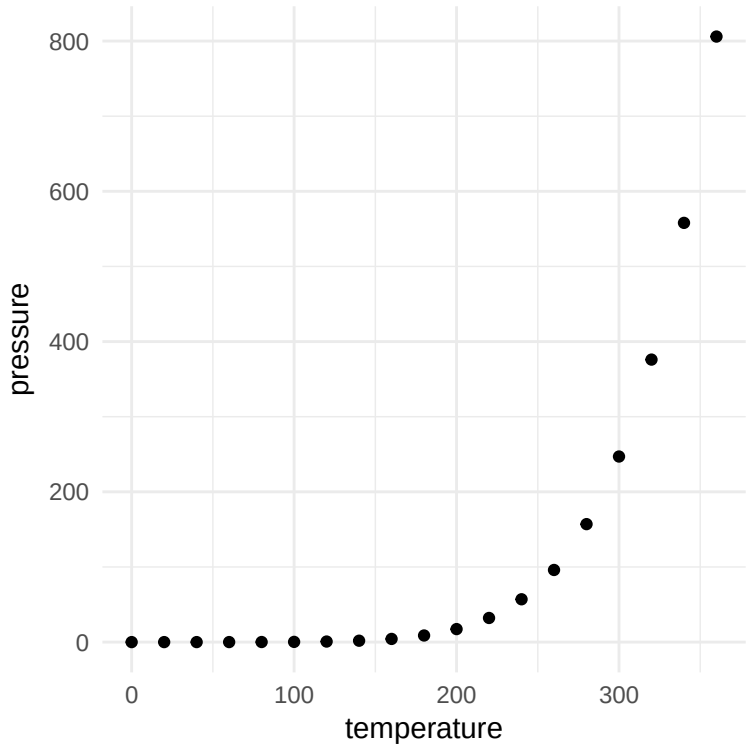

Figure 7.1: test a caption

$$\sum_{i}^{x} 5 + 1 \tag{7.1}$$

$$\sum_{i}^{x} 5 + 5 \tag{7.2}$$

7.1.2 Level 4

Theorem 7.1. *Here is A theorem.*

7.2 Including Plots

You can also embed plots, for example:

Note that the `echo = FALSE` parameter was added to the code chunk 7.1 to prevent printing of the R code that generated the plot. Bishop (2006)

also, lets put this in here

$$
\begin{aligned}
\text{Var}(\hat{\beta}) &= \text{Var}((X'X)^{-1}X'y) \\
&= (X'X)^{-1}X'\text{Var}(y)((X'X)^{-1}X')' \\
&= (X'X)^{-1}X'\text{Var}(y)X(X'X)^{-1} \\
&= (X'X)^{-1}X'\sigma^2 IX(X'X)^{-1} \\
&= (X'X)^{-1}\sigma^2
\end{aligned} \tag{7.3}
$$

and now lets referene (7.3) for god sake Note that the `echo = FALSE` parameter was added to the code chunk 7.1 to prevent printing of the R code that generated the plot.

Note that the `echo = FALSE` parameter was added to the code chunk 7.1 to prevent printing of the R code that generated the plot.

Note that the `echo = FALSE` parameter was added to the code chunk 7.1 to prevent printing of the R code that generated the plot. Note that the `echo = FALSE` parameter was added to the code chunk 7.1 to prevent printing of the R code that generated the plot.

7.3 level chapeter

fdsfdsf sdf fsd f sdf sd f sf

8 let's try an ew cahpeter

$$
\begin{aligned}
\text{Var}(\hat{\beta}) &= \text{Var}((X'X)^{-1}X'y) \\
&= (X'X)^{-1}X'\text{Var}(y)((X'X)^{-1}X')' \\
&= (X'X)^{-1}X'\text{Var}(y)X(X'X)^{-1} \\
&= (X'X)^{-1}X'\sigma^2 IX(X'X)^{-1} \\
&= (X'X)^{-1}\sigma^2
\end{aligned} \tag{8.1}
$$

trterter (8.1)

References

Bishop, Christopher M. 2006. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Secaucus, NJ, USA: Springer-Verlag New York, Inc.

Hastie, Trevor, Robert Tibshirani, and Jerome Friedman. 2001. *The Elements of Statistical Learning*. Springer Series in Statistics. New York, NY, USA: Springer New York Inc.

James, Gareth, Daniela Witten, Trevor Hastie, and Robert Tibshirani. 2014. *An Introduction to Statistical Learning: With Applications in R*. Springer Publishing Company, Incorporated.

Kopf, Dan. 2015. "The Discovery of Statistical Regression." *Priceonomics*. https://priceonomics.com/the-discovery-of-statistical-regression/.

Legendre, Adrien-Marie. 1805. *Nouvelles Méthodes Pour La Détermination Des Orbites Des Comètes*. Firmin Didot, Paris, France.