# Regression Notes

**Ryan Gehring**

# Contents


**Theory** **3**

- Model Parametrization . . . 3
- Loss Function . . . 3
- Normal Equation . . . 3
- Coefficient Estimates . . . 3

**Application** **3**

- Java Implementation . . . 3
- Estimation of Bear Age from Body Measurements . . . 4

# Theory

## Model Parametrization

A vector of response variables $Y$ is estimated by the product of a matrix of predictive variables $X$ and coefficient vector $\beta$ with normal errors $\epsilon$.

$$Y = X\beta + \epsilon \tag{1}$$

## Loss Function

Minimize the sum of squared errors.

$$SSE = \epsilon^T \epsilon \tag{2}$$

$$SSE = (Y - X\beta)^T (Y - X\beta)$$

## Normal Equation

To minimize loss, differentiate with respect to the coefficients, set to zero:

$$0 = -2X^T (Y - X\beta) \tag{3}$$

## Coefficient Estimates

Multiply by $(X^T X)^{-1}$ (and absorb 2 into $\beta$) to get the least squares estimate $\beta$:

$$\beta = (X^T X)^{-1} X^T Y \tag{4}$$

# Application

## Java Implementation

See below for an implementation using Apache commons math. The Github Repo contains code to run the model against an arbitrary csv data file.

Listing 1: Regression using commons math

```
import org.apache.commons.math3.linear.* ;


class RegressionTools {

  public static RealMatrix solveForCoefficients(RealMatrix x, RealMatrix y) {
    return (new LUDecomposition(    x.transpose().multiply(x)      )
        .getSolver()
        .getInverse())
      .multiply(x.transpose())
      .multiply(y)   ;
  }

}
```

## Estimation of Bear Age from Body Measurements

Above code was used to generate predictions of bear ages from a variety of body measurements.
Bear Data (truncated for readability):

| AGE | SEX | HEADLEN | HEADWTH | NECK | LENGTH | CHEST | WEIGHT |
|---|---|---|---|---|---|---|---|
| 19 | 0 | 11 | 5.5 | 16 | 53 | 26 | 80 |
| 55 | 0 | 16.5 | 9 | 28 | 67.5 | 45 | 344 |
| 81 | 0 | 15.5 | 8 | 31 | 72 | 54 | 416 |
| 115 | 0 | 17 | 10 | 31.5 | 72 | 49 | 348 |
| 104 | 1 | 15.5 | 6.5 | 22 | 62 | 35 | 166 |
| 100 | 1 | 13 | 7 | 21 | 70 | 41 | 220 |
| 56 | 0 | 15 | 7.5 | 26.5 | 73.5 | 41 | 262 |
| 51 | 0 | 13.5 | 8 | 27 | 68.5 | 49 | 360 |
| 57 | 1 | 13.5 | 7 | 20 | 64 | 38 | 204 |
| 53 | 1 | 12.5 | 6 | 18 | 58 | 31 | 144 |
| 68 | 0 | 16 | 9 | 29 | 73 | 44 | 332 |
| 8 | 0 | 9 | 4.5 | 13 | 37 | 19 | 34 |
| 44 | 1 | 12.5 | 4.5 | 10.5 | 63 | 32 | 140 |

$R^2$ was measured to be 0.64.

| | |
|---|---|
| SEX: | 26.148987054376104 |
| HEADLEN: | 3.0528248476353443 |
| HEADWTH: | 3.791600738790277 |
| NECK: | 1.366288821513322 |
| LENGTH: | -0.11350398628450656 |
| CHEST: | -2.9977072569525833 |
| WEIGHT: | 0.3122111457892039 |