Chapter 17: Metric Predicted Variable with One Metric Predictor

Shubham Gupta

December 12, 2019

1 Introduction

- Scenarios such as predicting weight from height for a person
- Predicted variable: metric
- Predictor variable: metric
- Relationship between y and x will be a linear model with distributed residual radomness in y i.e *simple linear regression*
- Generalize linear regression in 3 ways
  - Use t distribution instead of normal distribution to accommodate outliers
  - Replace linear trend with quadratic trend
  - Hierachical model to determine individual trend and estimate group level trends as well

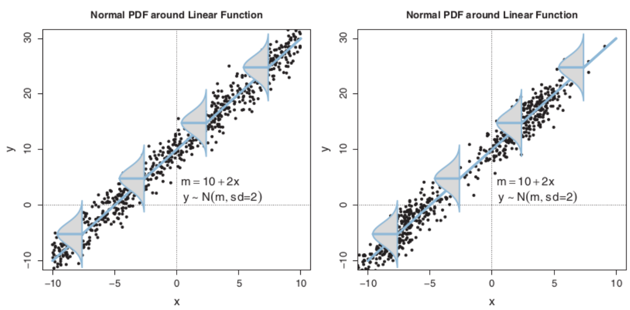

Figure 1: Points for a normally distributed function

- Function: $\mu = \beta_0 + \beta_1 x$
- Current model only specifies dependency of y on x. It does not show how x is generated and no prob dist assumed for x.
- **Homogeneity of Variance**: For every value of x, the variance in y is te same.

2 Robust Linear Regression

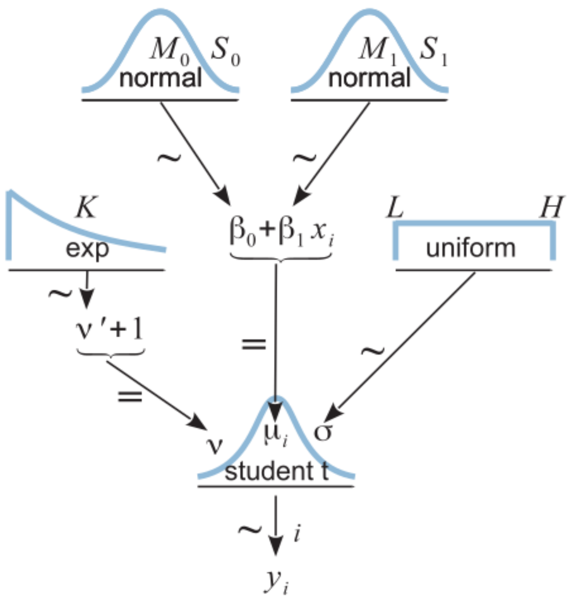

Figure 2: Robust Linear regression

- y is a t-distribution with mean μ
- μ has β_0 and β_1 which are both normal distributions
- Scale σ is a uniform prior
- Normality ν is an exponential prior

2.1 Standardizing data

- As shown in the figure, there are points where there are many regression lines flowing through them.
- These points are a collection of a large number of regression lines.
- Sampling from this tightly corelated space can be difficult.
- Two ways to make sampling faster:
  - Change sampling algo. Instead of Gibbs, use HMC.
  - Transform regression lines to ensure no strong corelation between slopes and intercepts.
- **Standardize**: Rescaling data relative to mean and SD.
- If input data is standardized, output will also be on a standardized scale.

2.2 Interpreting posterior distribution

- Model estiamtes are tighter for example with 300 samples.
- The graph suggests that there might be positive skew in the dataset.

3 Hierachical regression on Individuals within groups

- We can estimate reression lines for each individual and group if we have data in the form $x_{i|j}l$ and $y_{i|j}$. $i|j$ represents i^{th} observation for j_{th} individual.
- Goal: Describe each individual with linear regression and estimate group level characteristics.

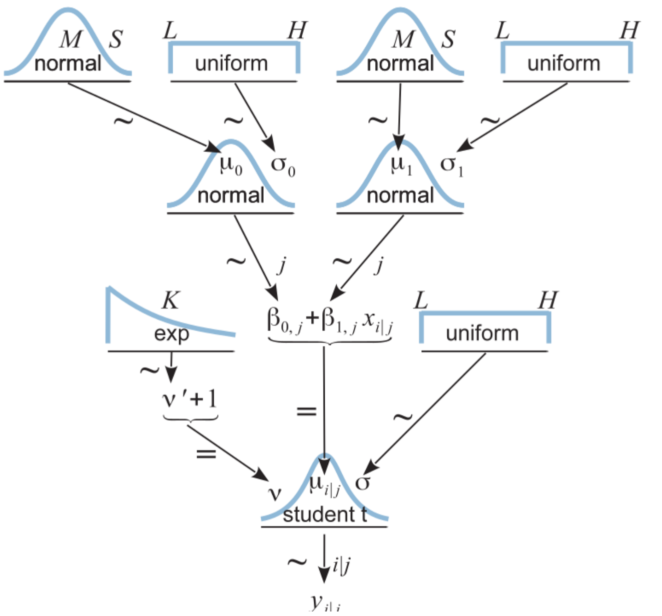

Figure 3: Robust Hierachical Linear Regression Model

- Mate this is new