1 Appendix

1.1 Source code

All the source code is located in this GitHub repository.

1.2 L_0 regression on matrix form

The cost function we use for OLS regression is the residual sum of squares (RSS) function:

$$\mathrm{RSS}(\beta) = \sum_{i=1}^{N} (y_i - \hat{y}_i)^2 \,.$$

Changing into matrix notation, we get

$$\mathrm{RSS}(\beta) = (\mathbf{y} - \hat{\mathbf{y}})^T(\mathbf{y} - \hat{\mathbf{y}}) = (\mathbf{y} - \mathbf{X}\beta)^T(\mathbf{y} - \mathbf{X}\beta),$$

which we can differentiate with respect to β to find the minimum.

$$\frac{\partial \mathrm{RSS}}{\partial \beta} = -2\mathbf{X}^T(\mathbf{y} - \mathbf{X}\beta).$$

Assuming full column rank for $\mathbf{X}$, $(\mathbf{X}^T\mathbf{X})$ is thus positive definite (and importantly, invertible). Setting the first derivative to 0, we get

$$\mathbf{X}^T(\mathbf{y} - \mathbf{X}\beta) = 0$$

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

1.3 Deriving the bias-variance decomposition

We a