Neural Networks and Multivariate Linear Regression

Josh Christensen

BYU

INTRODUCTION

BYU

Objectives

- Generally compare point predictions from Neural Networks and Multivariate Linear Regression
- Assess how covariance in a multivariate response affects predictions
- Assess how non-linearity affects predictions

BYU

Why these two and point predictions?

- Linear regression is the default
- Neural networks are often presented as the superior alternative
- Point estimates is the primary area where the two sets of use cases overlap

BYU

DATA

Data Set Information

- University of California Irvine Machine Learning Repository
- Characterization of over 200 automobiles in terms of structure and function
- Insurance information:
  - Symboling
  - Normalized losses (not used due to missing values)

BYU

Variables

1. Symboling
2. Make
3. Fuel Type
4. Aspiration
5. Number of Doors
6. Body Style
7. Drive wheels
8. Engine Location
9. Wheel Base
10. Length
11. Width
12. Height
13. Curb Weight
14. Engine Type
15. Number of Cylinders
16. Engine Size
17. Fuel System
18. Bore
19. Stroke
20. Compression Ratio
21. Peak RPM
22. Price

Response Variables:

1. City MPG
2. Highway MPG
3. Horsepower

BYU

METHODS

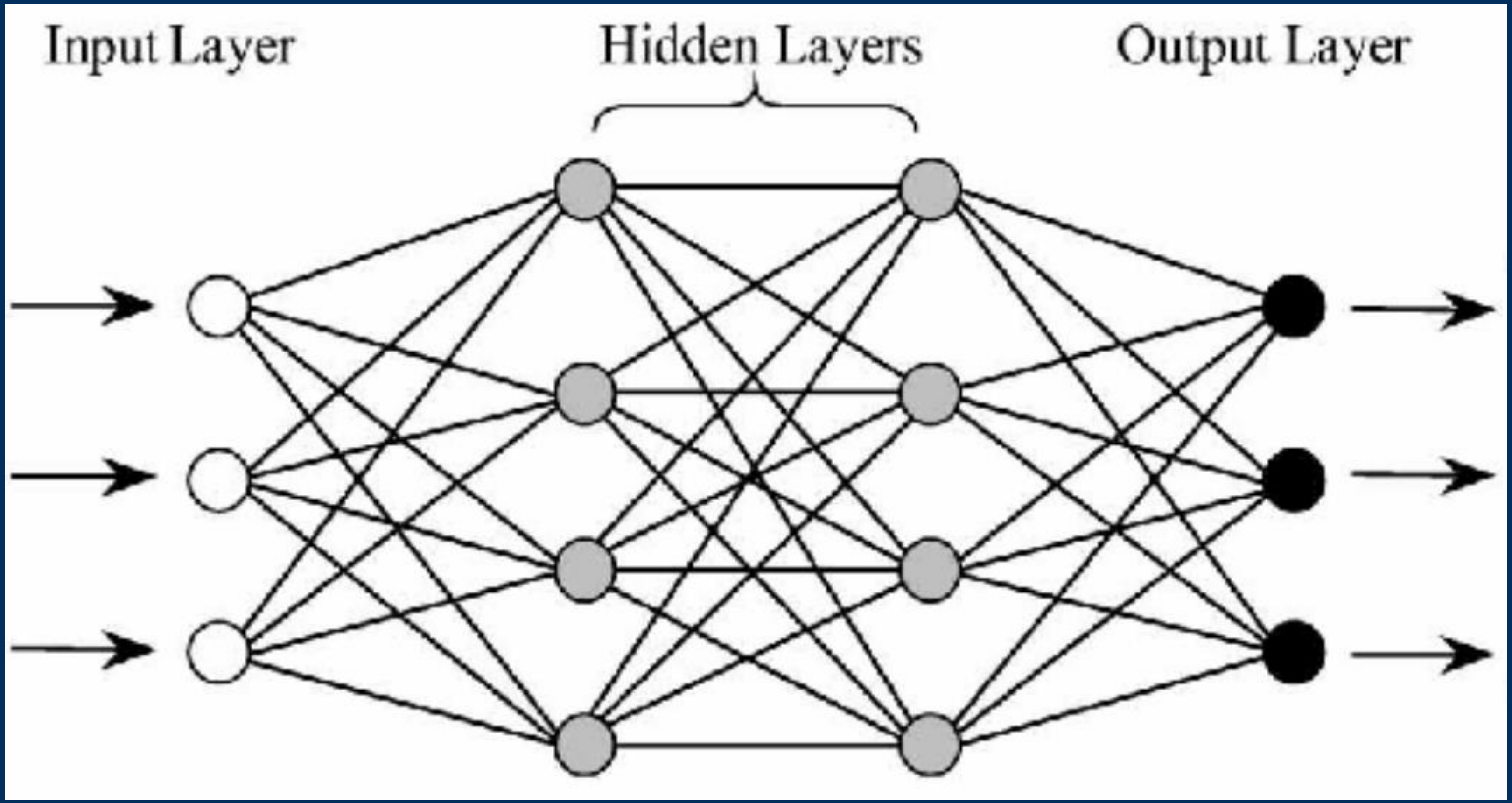
BYU

Multivariate Regression

- $\boldsymbol{Y} = \boldsymbol{XB} + \boldsymbol{\Xi}$
- For point predictions we only need an estimate for $\widehat{\boldsymbol{B}}$ since our point estimate is $E(\mathbf{Y}) = \mathbf{XB}$
- $\widehat{\boldsymbol{B}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}$
- $\widehat{\boldsymbol{y}} = \boldsymbol{x}\widehat{\boldsymbol{B}}$
- Since $\widehat{\boldsymbol{y}}$ is just a linear combination of $\boldsymbol{x}$ it cannot capture non-linear effects unless we explicitly include them

BYU

Neural Networks



BYU

Neural Networks

- Linear function

$$\mathrm{o} = \boldsymbol{c}'\boldsymbol{w} + bias$$

- Activation function (ReLU)

$$f(x) = \begin{cases} x & x \geq 0 \\ 0 & x < 0 \end{cases}$$

- Backpropagation
    - Iterative gradient based optimization
    - Usually paired with stochastic gradient descent

BYU

Simulation Studies

Simulation Study 1

- 10 covariates with values drawn from standard normals
- Beta values
  - $\beta_1 = (1, 2, \dots, 10)'$
  - $\beta_2 = (0.1, 0.2, \dots, 1)'$
  - $\beta_3 = (-0.5, -0.4, \dots, 0.4)'$
- Standard deviations
  - (1,10,5)

BYU

Simulation Study 1

- Run 1
  - Independent responses
- Run 2
  - $\rho_1 = (0,0,0)$
  - $\rho_2 = (0.9, -0.1, 0.3)$
- Run 3
  - $\rho_1 = (0.8, -0.1, -0.5)$
  - $\rho_2 = (-0.8, 0.1, 0.5)$

BYU

Simulation Study 1

	Run 1	Run 2	Run 3
Multivariate Regression	4.282	4.275	4.271
Neural Network	4.256	4.256	4.249

Table 1: Mean absolute error results from the first simulation study.

BYU

Simulation Study 2

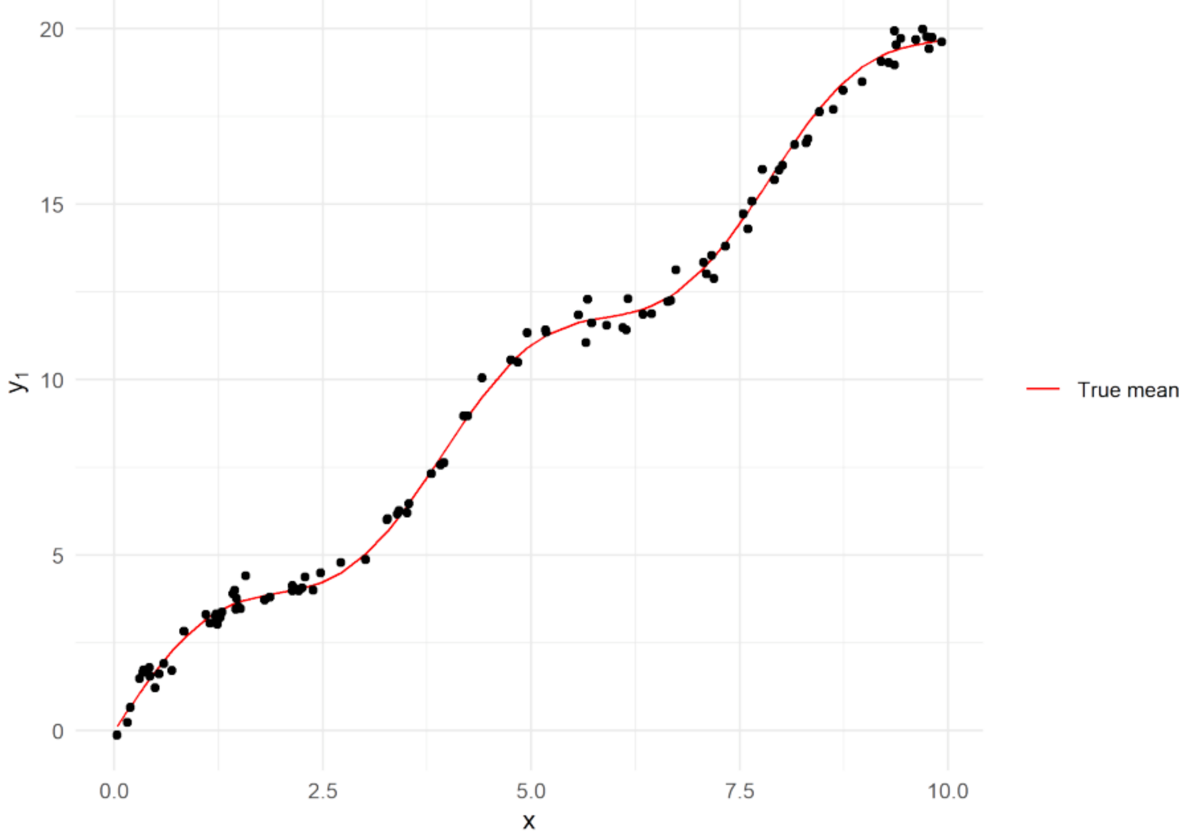


Figure 2: Demonstration of the data generating process for simulation two.

BYU

Simulation Study 2

	σ^2							
	Multivariate Regression				Neural Network			
	0.1	1	3	8	0.1	1	3	8
$\rho = 0.1$	0.654	0.992	1.499	2.372	0.554	0.964	1.500	2.387
$\rho = 0.3$	0.658	0.985	1.501	2.333	0.545	0.948	1.505	2.366
$\rho = 0.5$	0.654	0.994	1.501	2.362	0.563	0.965	1.500	2.405
$\rho = 0.7$	0.658	1.009	1.535	2.410	0.545	0.964	1.547	2.428
$\rho = 0.9$	0.645	0.979	1.570	2.310	0.548	0.946	1.566	2.360

Table 2: Mean absolute error results from the second simulation study.

Results

BYU

Car data

- Included all variables
- Two layer neural network with 40 nodes per layer
- Out of sample MAE
  - Neural Network: 3.86
  - Multivariate Regression: 35.57

Conclusion

BYU

Practical Recommendations

- Generally don’t use neural nets for regression
- Use cases:
  - Unstructured data
  - Non-traditional data
  - Difficult to explore non-linearities or collinearity with curves that trees may struggle with
- Consider whether prediction is REALLY your ONLY goal
  - Consider “tuning” your multivariate regression using validation/test sets

Future Work

- Investigate theoretical bases and effects of using validation sets for tuning and test sets for evaluation
- Investigate the effects of tuning parameters on neural network regression predictions
    - Activation functions
    - Optimization routine
    - Learning rate
    - Early stopping rules
    - Pruning rules

BYU