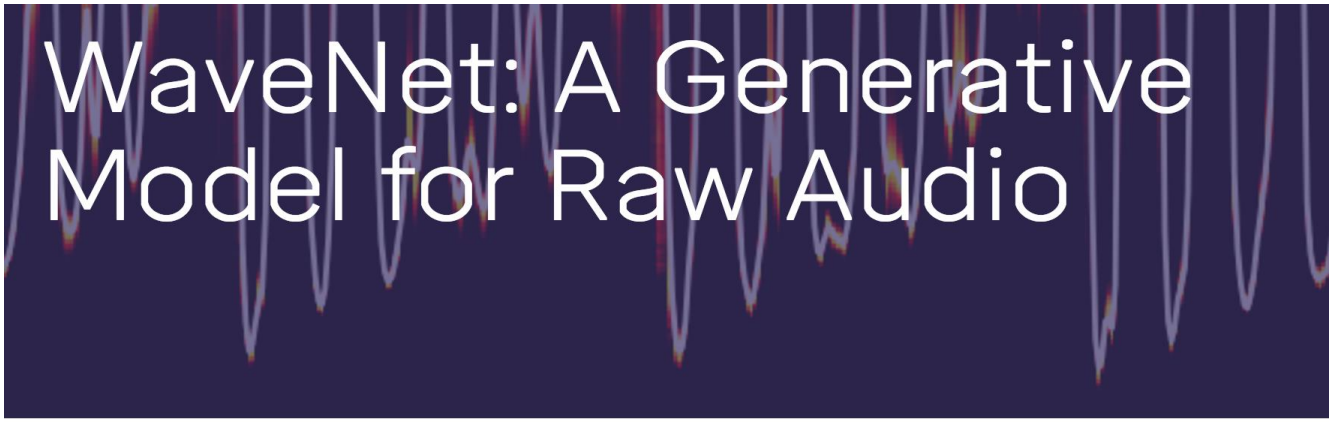
Machine Learning applied to Planetary Sciences

PTYS 595B/495B

Leon Palafox

https://leonpalafox.github.io/MLClass/

News of the day



This post presents **WaveNet**, a deep generative model of raw audio waveforms. We show that WaveNets are able to generate speech which mimics any human voice and which sounds more natural than the best existing Text-to-Speech systems, reducing the gap with human performance by over 50%.

https://deepmind.com/blog/wavenet-generative-model-raw-audio/

Last Class Recap – Logistic Regression

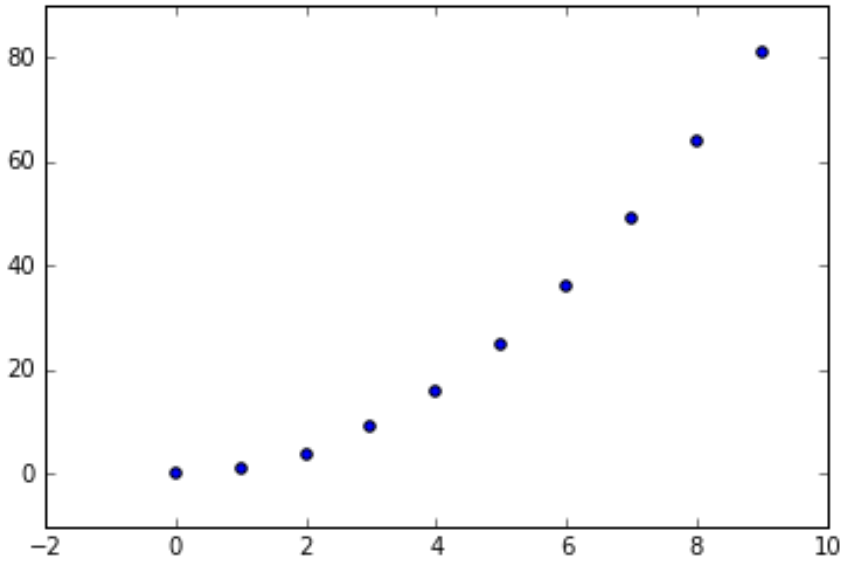
- Imagine you want to sell you car:
  - How much do you ask for it:
    - Mileage
    - Year
    - Color
    - Options
    - Condition

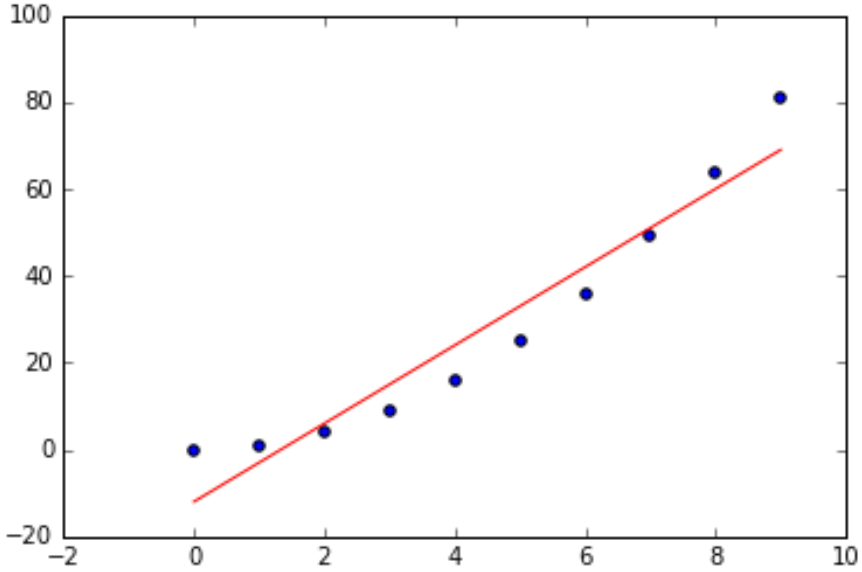
Matrix form

$$\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \quad \mathbf{X} = \begin{pmatrix} \mathbf{x}_1^{\mathrm{T}} \\ \mathbf{x}_2^{\mathrm{T}} \\ \vdots \\ \mathbf{x}_n^{\mathrm{T}} \end{pmatrix} = \begin{pmatrix} x_{11} & \cdots & x_{1p} \\ x_{21} & \cdots & x_{2p} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{np} \end{pmatrix},$$

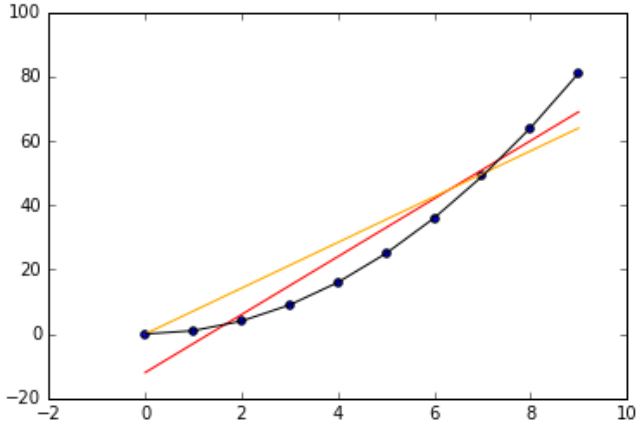
$$\beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{pmatrix} \qquad \mathbf{y} = \mathbf{X}\beta$$

Feature Engineering





By adding an extra squared factor, we got a perfect fit

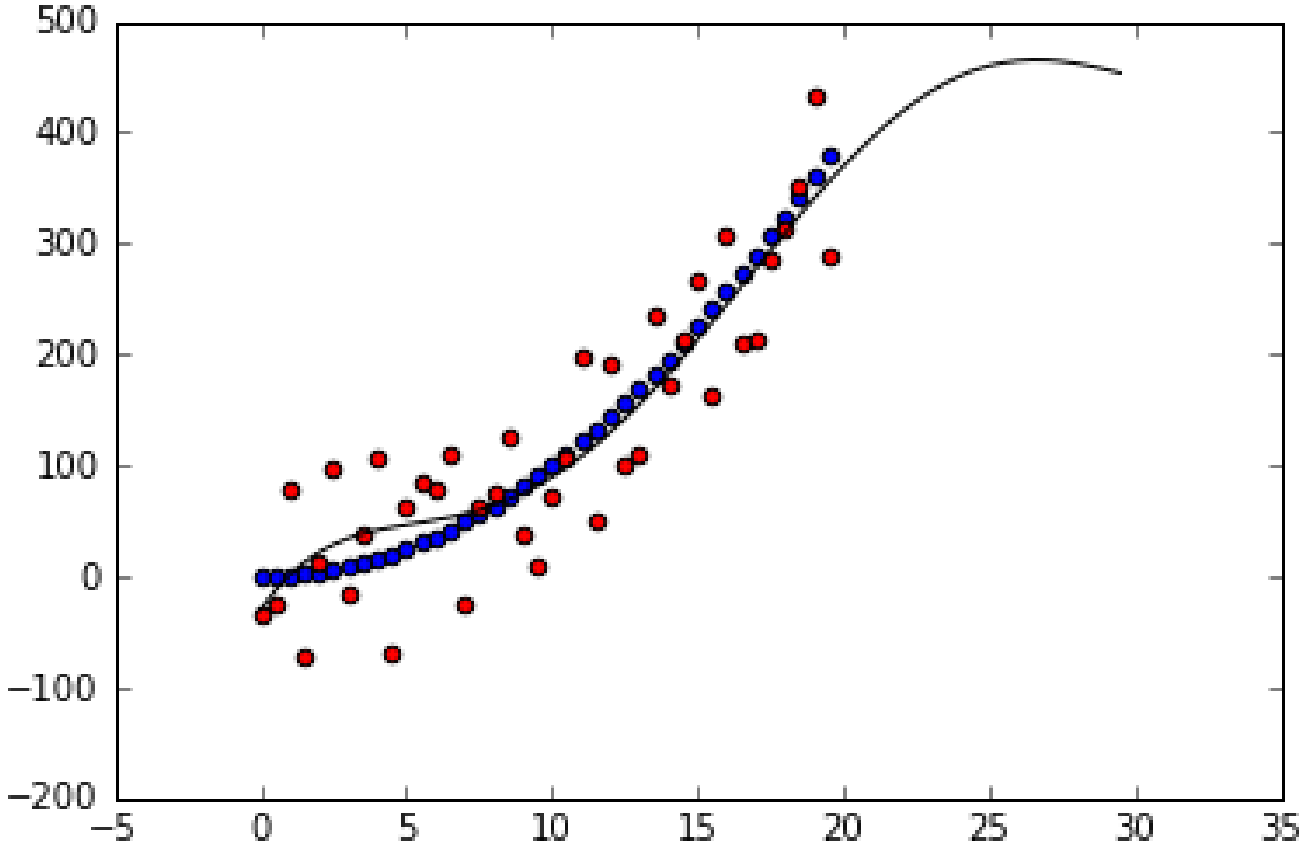


What is the difference

- Large amount of data:
  - Gradient descent rules!
  - Matrix Algebra is a pain.
- Small datasets
  - Matrix algebra is way faster.

- In practice, most implementations use Gradient Descent, since datasets get large very fast.
  - Vowpal Wabbit (http://hunch.net/~vw/)

How to control that polynomial problem?



Logistic Regression – Our second tool (hammer)

- Classification
  - Discriminate between two or more classes.
  - Wildly popular, and the main ML undertaking.
- Student Performance
  - Success or Failure
    - Demographics
    - School District
    - Parent’s work

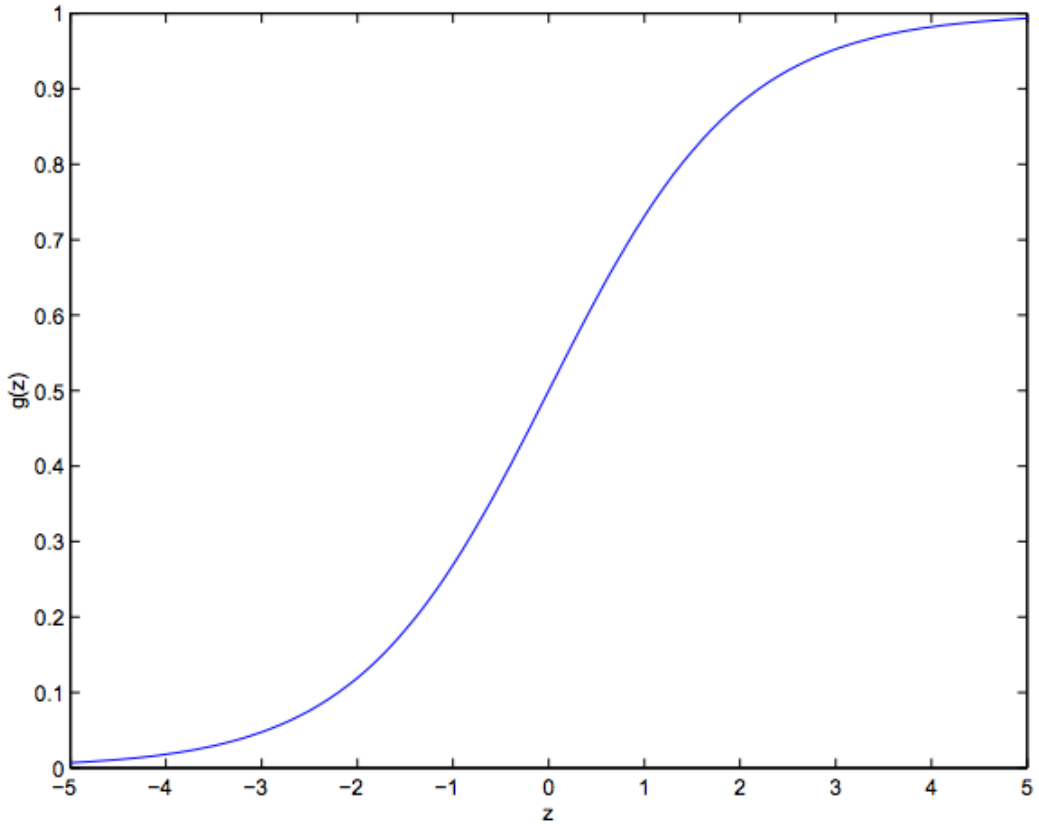
Logistic regression

$$h_\beta(x) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

We need to find a way to make clamp real numbers between two values.

- Set a pivot value (less than or higher than) (ReLU)
- Set a function f(x) capable of clamping the value.

Sigmoid function



$$g(z) = \frac{1}{1 + e^{-z}}$$

Sigmoid function

$$h_\beta(x) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

$$g(\beta^T x) = \frac{1}{1 + e^{-\beta^T x}}$$

$$\boxed{h_\beta(x) = g(\beta^T x)}$$

Optimization

- Logistic regression doesn’t have a nice closed matrix form.
- And we can’t use the Linear Regression cost, because the objective values are either 1 or 0.
- Let’s think about the cost function

Optimization

$$p(y = 1|x; \beta) = h_\beta(x)$$

$$p(y = 0|x; \beta) = 1 - h_\beta(x)$$

$$p(y|x; \beta) = h_\beta(x)^{1-y}(1 - h_\beta(x))^y$$

Optimization

$$L(\beta) = p(y|x;\beta)$$

$$L(\beta) = \prod_{i=1}^{m} p(y^i|x^i;\beta)$$

$$l(\beta) = log(L(\beta))$$

Logistic Regression Regularization

- We can also add penalties to the final loss, to prevent overfitting.

- Everything we have been doing with linear regression, can be done with logistic regression.
  - Add penalty.
  - Add L1 Norm.