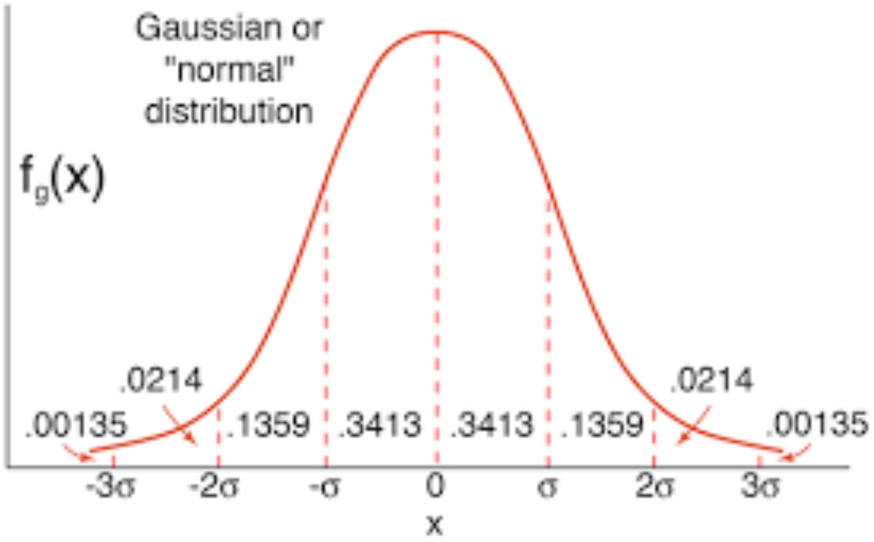
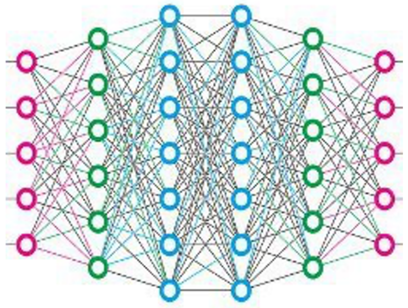
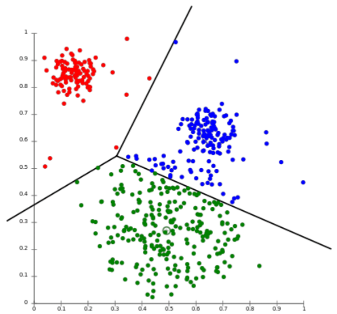
CS 445
Introduction to Machine Learning

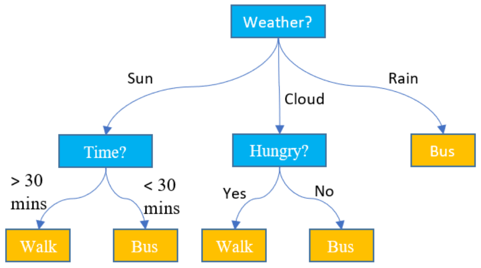
Normal Isn't Everything



Instructor: Dr. Kevin Molloy







PA 1 Review

44	**18**	**8**	**5**
23	31	8	13
12	16	21	26
9	11	11	44

Plan for Today

Last time:

- Naïve Bayes as a Classifier

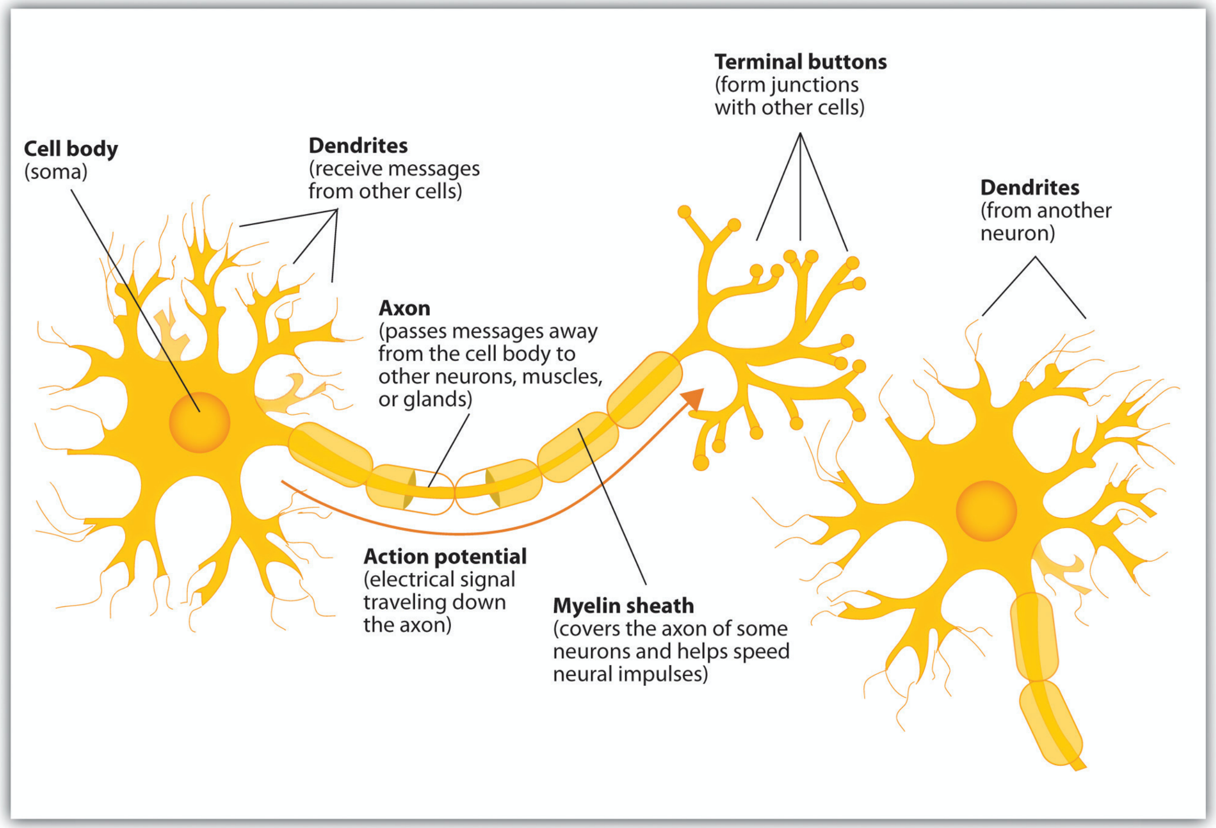
Today:

Start our Discussion on Neural Networks

Another Approach

Neurons

- Neurons communicate using discrete electrical signals called "spikes" (or action potentials)
- "Spikes" travel along axons, and reach terminals, where neurotransmitters are released.
- Postsynaptic neurons respond by allowing current to flow in (or out).
- If voltage crosses a threshold a spike is created.



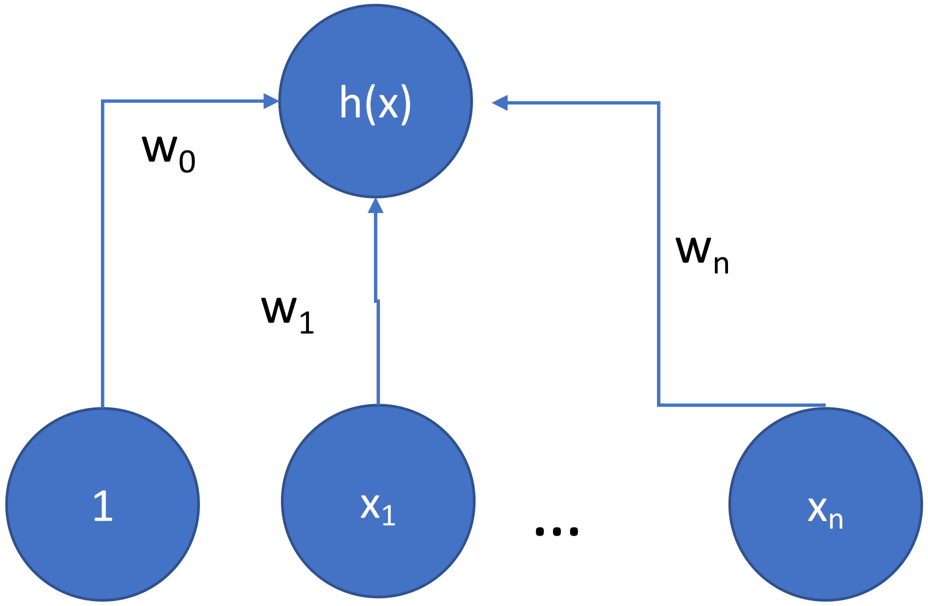
Multivariate Linear Regression

Multi-dimensional

$$h(x_1, x_2, \ldots x_n) = w_0 + w_1 x_1 + \ldots + w_n x_n$$

OR

$$h(x) = w^T x$$



Linear Regression – The Neural View

Multi-dimensional

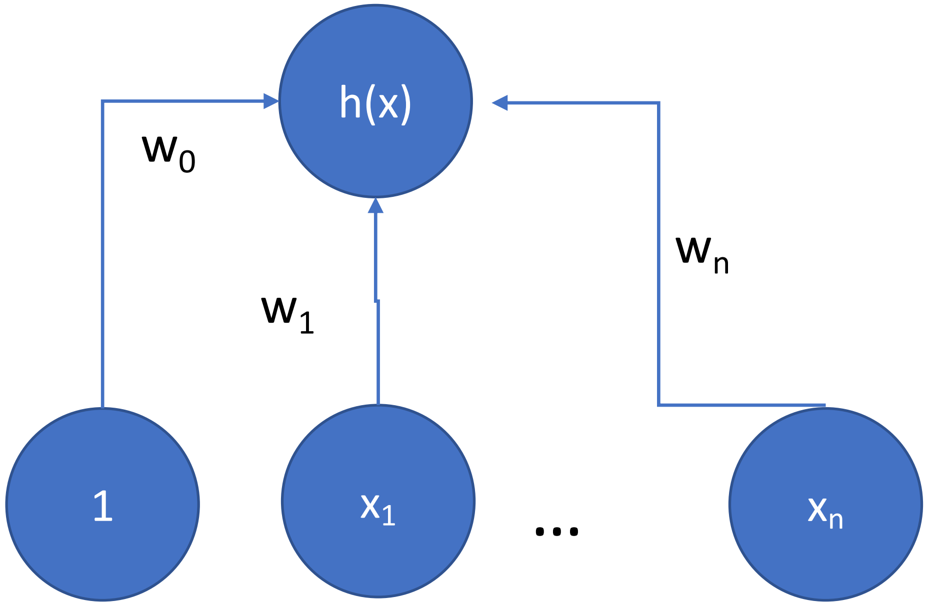
Input: x

Desired output: y

Weight = w

$$h(x) = w^T x$$

- Given a set of inputs and a corresponding desired output, we need to select *w*.
- What is going on here?



Gradient Descent

Gradient Descent

One Approach

1. Take the derivative of the function f'(w)
2. Guess a value for w:
3. Move a little bit according to the derivative. $\hat{w} = \hat{w} - \eta f'(\hat{w})$

Partial Derivatives

- Derivative of a function of multiple variables, with all but the variable of interest held constant.

$$f(x,y) = x^2 + xy^2$$

$$f_x(x,y) = 2\text{x} + y^2$$

OR

$$\frac{\partial f(x,y)}{\partial x} = 2\text{x} + y^2$$

$$f_y(x,y) = 2\text{xy}$$

OR

$$\frac{\partial f(x,y)}{\partial y} = 2\text{xy}$$

Gradient

- The gradient is just the generalization of the derivative to multiple dimensions.

$$\nabla f(\mathbf{w}) = \begin{bmatrix} \frac{\partial f(\mathbf{w})}{\partial w_1} \\ \frac{\partial f(\mathbf{w})}{\partial w_2} \\ \vdots \\ \frac{\partial f(\mathbf{w})}{\partial w_n} \end{bmatrix}$$

- Gradient descent update:

$$\hat{\mathbf{w}} \leftarrow \hat{\mathbf{w}} - \eta \nabla f(\hat{\mathbf{w}})$$

Gradient Descent for MVLR

- Error for the multi-dimensional case:

$$Error_E(\mathbf{w}) = \sum_{e \in E} \frac{1}{2}\left(y_e - \mathbf{w}^{\mathbf{T}}\mathbf{x}_{\mathbf{e}}\right)^2$$

$$\frac{\partial Error_E(\mathbf{w})}{\partial w_i} = \sum_{e \in E}\left(y_e - \mathbf{w}^{\mathbf{T}}\mathbf{x}_{\mathbf{e}}\right)\left(-x_{e,i}\right)$$

$$= -\sum_{e \in E}\left(y_e - \mathbf{w}^{\mathbf{T}}\mathbf{x}\right)x_{e,i}$$

- The new update rule:

$$w_i \leftarrow w_i + \eta \sum_{e \in E}\left(y_e - \mathbf{w}^{\mathbf{T}}\mathbf{x}\right)x_{e,i}$$

- Vector version:

$$\mathbf{w} \leftarrow \mathbf{w} + \eta \sum_{e \in E}\left(y_e - \mathbf{w}^{\mathbf{T}}\mathbf{x}\right)\mathbf{x}_{\mathbf{e}}$$

Analytical Solution

$$\mathbf{w} = (X^T X)^{-1} X^T y$$

- Where X is a matrix with one input per row, y the vector of target values.

Lines