My basic AI notebook (tips that I think matter)

M. Yas. Davoodeh

September 27, 2020

1 Math

A vector is a $x \times 1$ matrix.

Vector $\vec{x}$ has a Weight property which indicates its travel distance (Euclidean). Weight of $\vec{x}$ is indicated by $\|\vec{x}\|$.

Euclidean or Normal distance is calculated like below where Ds is the number of dimensions:

$$d(p, q) = d(q, p) = \sqrt{\sum_{i=1}^{Ds} (q_i - p_i)^2} \tag{1}$$

2 Not-Math

2.1 Hypothesis

$h_\Theta(x)$ is the example function for Θ parameters as in $h_\Theta(x) = \Theta^T x$. A good h gives a correct answer y for any given x.

2.2 Cost function

$J(x)$ is the difference between a guessed answer (h_Θ) and the correct one. In the simplest form:

$$J(x) = h_\Theta(x) - y \tag{2}$$

Linear–Regression J for a set like $x = \{(x_1, y_1), (x_2, y_2), \ldots, (x_m, y_m)\}$:

$$J(x) = \frac{\sum_{i=1}^{m}(h_\Theta(x_i) - y_i)^2}{2m} \tag{3}$$

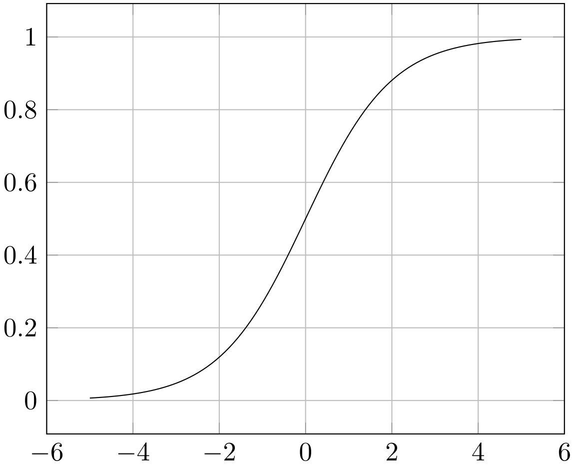
The power 2 prevents negative values from cancelling out the positive ones. Absolute function ($|x|$) can be used as well. The reason that the power 2 function is more used is that it penalizes outlier values exponentially.

2.3 Logistic/Sigmoid function

$$g(x) = \frac{L}{1 + e^{-k(x-x_0)}} \tag{4}$$

Where

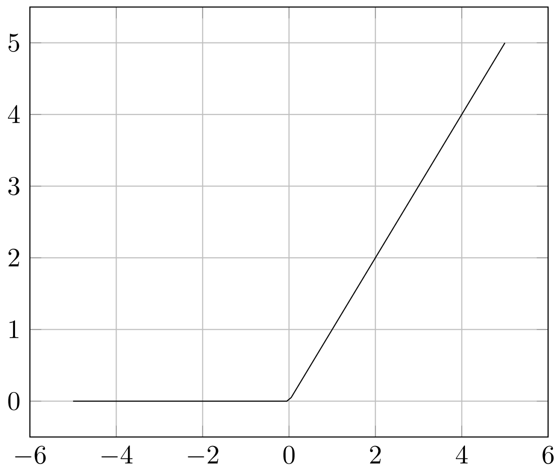
x_0 = the x value of the sigmoid's midpoint (default=0),
L = the curve's maximum value (default=1),
k = the logistic growth rate or steepness of the curve (default=1).



2.4 ReLU

Rectified Linear Unit

$$\sigma(x) = R(x) = max(0, x) \tag{5}$$



3 Normalization

Normalization aims to reduce processing required or simplifying values.

$$s = \text{max}(x) - \text{min}(x) \quad \text{OR} \quad s = \underbrace{\sigma(x)}_{\text{std}} \tag{6}$$

3.1 Feature Scaling

By dividing a constant, s, to x in order to make the change range closer to $-1 \leq x \leq 1$, we scale the features.

$$x' = \frac{x}{s} \tag{7}$$

To rescale a range between an arbitrary set of values $[a, b]$, the formula becomes:

$$x' = \frac{(x - \text{min}(x))(b - a)}{s} \tag{8}$$

3.2 Mean Noramlization

$$x' = \frac{x - \overbrace{\mu}^{\bar{x}}}{s} \tag{9}$$

3.3 Making decision based on decision boundary

For any given decision boundary $\vec{d}$, there is a perpendicular vector, $\vec{w}$, to it from the origin $((0,0))$.

To determine if unknown vector $\vec{u}$ if over $\vec{d}$ or not we check if $\vec{u} \cdot \vec{w} + bias > 0$ or not.

If $\vec{u} \cdot \vec{w} + bias > 0$, $\vec{u}$ is over $\vec{d}$.
else If $\vec{u} \cdot \vec{w} + bias < 0$, $\vec{u}$ is not over $\vec{d}$.
else If $\vec{u} \cdot \vec{w} + bias = 0$, $\vec{u}$ is on the decision boundary $\vec{d}$.