Summary

Luca Bracone

March 17, 2020

1 Basic Machine Learning

1.1 Neural Networks

Definition 1.1 (Perceptron) *A Perceptron is a function $p : \mathbb{R}^n \to \mathbb{R}$ given by*

$$p(x_1, \ldots, x_n) = g\left(w_0 + \sum_{i=1}^{n} w_i x_i\right)$$

Where $w_1, \ldots, w_n \in \mathbb{R}$ are called the weights of p, w_0 is called the bias, and $g : \mathbb{R} \to \mathbb{R}$ is any non-linear function called the activation. Most of the time

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

the sigmoid function.

Definition 1.2 (Artificial Neural Network) *A (forward-feeding) Artificial Neural Network (ANN) is a function $T : \mathbb{R}^n \to \mathbb{R}^m$ that can be decomposed as such*

$$T = l_k \circ \cdots \circ l_1$$

Where the $l_i : \mathbb{R}^{n_j} \to \mathbb{R}^{n_{j+1}}$ are called layers and are essentially just vectors of perceptrons

$$l_i(\vec{x}) = (p_{i,1}(\vec{x}), \ldots, p_{i,n_j}(\vec{x}))$$

. In the context of neural networks perceptrons are also called nodes.

Definition 1.3 (Loss Function) *For some ANN, T and some input $(x_1, ..., x_n)$, we have a desired output $(y_1, ..., y_n)$ we would like T to match. Write $(\hat{y}_1, ..., \hat{y}_m)$ the output T actually produced. The loss (or sometimes, cost) of T for this input is a function that measures the "distance" between the produced and desired output. For most basic applications,*

$$\mathcal{L}(W; x) = \sum_{i=1}^{m} ||y_i - \hat{y}_i||_2^2$$

Definition 1.4 (Empirical Loss) *For a set of inputs X, and T a NN, the Empirical Loss is*

$$J(W) = \frac{1}{|X|} \sum_{\vec{x} \in X} \mathcal{L}(W; x)$$

the average loss over X.

1.2 Gradient Descent

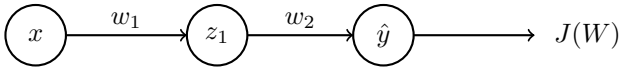
Any neural network is defined by the weights of its perceptrons. Therefore, in order to have the neural network behave as desired (namely being as close as possible to the desired outputs) we need to find an efficient way to tweak the weights such that the empirical loss becomes as low as possible, when tested on some new inputs. Applying this algorithm repeatedly is called "training".

1.2.1 Gradient Descent

Essentially it involves taking a step in the "right" direction until we reach a stopping condition, most of the time it's when the steps we are taking are too short. In other words update the weights $W \leftarrow W - \eta \nabla J$. Where η is a constant that has to be well chosen. If too small, gradient descent stops when it meets the slightest uphill, and thus misses a better minimum which might be close by. If too big, gradient descent can only take giant steps and diverges to infinity.

1.2.2 Gradient Backpropagation

In practice, calculating the derivative to some particular weight is quite simple, consider the following neural net:



It's just an application of the chain rule:

$$\frac{\partial J}{\partial w_2} = \frac{\partial J}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial w_2}$$
$$\frac{\partial J}{\partial w_1} = \frac{\partial J}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial z_1} \frac{\partial z_1}{\partial w_1}$$

1.2.3 Data Batching

Since calculating ∇J is computationally intensive, consider taking a random subset of your data, and applying gradient descent to it instead. the convergence is going to be more erratic, depending on the size of the subset chosen, but the time gained makes it worth.

1.3 Overfitting

When training our neural network, it might happen that it essentially "learns the dataset by heart" making it unable to properly tackle on never seen before data points. There are a few methods to avoid such a problem:

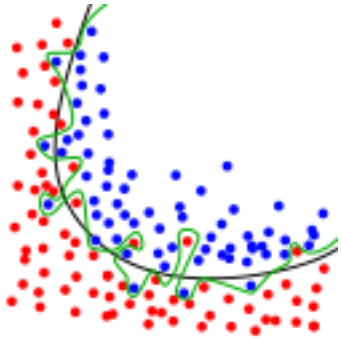


Figure 1: A parabola+noise is being misunderstood as a complicated winding line [1]

[1] Courtesy of wikimedia foundation https://commons.wikimedia.org/wiki/File:Overfitting.svg

1.3.1 Dropout

This technique involves randomly setting some nodes to zero during training. This makes the network not rely too much on some path.

1.3.2 Early Stopping

After each application of gradient descent, test the network. If it performs worse than it did previously, stop the training process.

2 Simplicial Complexes

Definition 2.1 (convex hull) *let $u_0, ..., u_k \in \mathbb{R}^n$ their convex hull is*

$$conv\{u_0, ..., u_k\} = \left\{ x = \sum_{i=0}^{n} \lambda_i u_i \quad s.t. \ \sum_{i=0}^{n} \lambda_i = 1; \ \lambda_i > 0 \ \forall i \right\} \qquad (\star)$$

Definition 2.2 (affinitely independent, k-simplex) *$u_0, ..., u_k \in \mathbb{R}^n$ are said to be affinitely independent if any point in their convex hull is written as in $(\star)$ uniquely. In that case instead of convex hull, we speak of k-simplex.*

Definition 2.3 (face, coface) *Let $\{u_{i_0}, ..., u_{i_m}\} \subseteq \{u_0, ..., u_k\}$ we say that $\tau = conv\{u_{i_1}, ..., u_{i_m}\}$ is a face of $\sigma = conv\{u_1, .., u_k\}$ which we write $\tau < \sigma$. Equivalently σ is a co-face of τ.*

Definition 2.4 (simplicial complex) *A simplicial complex K is a collection of simplices such that:*

(S1) $\forall \sigma \in K, \ \tau < \sigma \implies \tau \in K$.

(S2) $\forall \sigma, \tau \in K, \ \sigma \cap \tau$ is either empty or a face.

The dimension of K is $\dim K = \max_{\sigma \in K} \dim \sigma$ Where the dimension of a k-simplex is k. The Underlying space of K, $|K|$ is the topological space $\bigcup_{\sigma \in K} \sigma$ with the subspace topology.

Definition 2.5 (triangulation) *Let X be a topological space and K a simplicial complex, if there exists a homeomorphism $\phi : X \to |K|$ we say that X is triangulable. In that case the couple (K, ϕ) is called a triangulation of X*

Definition 2.6 (sub-complex) *A subset $L \subseteq K$ is a subcomplex of K, if it is itself a complex. It is called full if it contains all vertices of K.*

Example 2.1 (skeleton) *The j-skeleton of complex K is $K^{(j)} = \{\sigma \in K \ s.t. \ \dim \sigma \leq j\} \subseteq K$. In particular the 0-skeleton is called $vert(K)$ the vertices of K*

Definition 2.7 (star and link) *Let K be a simplicial complex and pick $\sigma \in K$. Its star, $st\sigma = \{\tau \in K \ s.t. \ \tau < \sigma\}$. unfortunately it's not a sub-complex of K because it's not closed under taking faces. For that we define the closed star $\overline{st}\sigma$ the smallest sub-complex that contains the star. Similarly, the link of a simplex $lk\sigma = \{\tau \in \overline{st}\sigma \ s.t. \ \tau \cap \sigma = \emptyset\}$ the set of simplices in the closed star that don't touch σ*

Definition 2.8 (simplicial maps, homeomorphisms) *Let K, L be two simplicial complexes and $\phi : vertK \to vertL$ a map such that vertices of every simplex in K gets mapped to a vertex of a simplex in L, in that case ϕ is called a vertex map. The map $f : |K| \to |L|$ defined by $\sum_{i=0}^{n} \lambda_i u_i \mapsto \sum_{i=0}^{n} \lambda_i \phi(u_i)$ is called a simplicial map. If ϕ is bijective and ϕ^{-1} is also a vertex map, we speak instead of simplicial homeomorphism.*

3 Simplicial Homology

Definition 3.1 (p-chain) *Let K be a simplicial complex p-chains are formal sums of p-simplices of K over a field:*

$$c = \sum_{i=1}^{n} a_i \sigma_i$$

componentwise addition makes it into a commutative group $(C_p, +)$. For the time being the field in question is $\mathbb{F}_2$.

Definition 3.2 (boundary) *the boundary of a p-simplex is the sum of its $(p-1)$-faces:*

$$\partial_p c = \sum_{j=1}^{n} [\sigma_1, \ldots, \cancel{\sigma_j}, \ldots, \sigma_p]$$

Turns out $\partial_p : C_p \to C_{p-1}$ is a group homomorphism. for a simplicial complex K we define its chain complex:

$$\mathcal{C}_K : \cdots \longrightarrow C_p \xrightarrow{\partial_p} C_{p-1} \xrightarrow{\partial_{p-1}} \cdots \xrightarrow{\partial_1} C_0$$

Lemma 3.1 (Fundamental of Homology) *For all p, $d \in C_{p+1}$, we have $\partial_p \partial_{p+1} d = 0$*

Whenever we have an interesting homomorphism it is natural to want to know more about its kernel and image, in this case they have special names

Definition 3.3 (p-boundaries and p-cycles) *p-cycles are* $\ker \partial_p = \{c \in C_p \text{ s.t. } \partial c = 0\} = Z_p$
p-boundaries are $Im\, \partial_{p+1} = \{c \in C_p \text{ s.t. } \exists d \in C_{p+1},\ c = \partial d\} = B_p$

Note that if $c \in B_p$ then $\partial c = \partial \partial d = 0$ so, $B_p \subseteq Z_p$. But in general $Z_p \neq B_p$ in particular we are interested in the p-cycles that are not p-boundaries.

Definition 3.4 (homology groups and Betti numbers) *for a complex K its p-th homology group is the quotient $H_p Z_p / B_p$.*
The p-th Betti number is β_p the number of generators of H_p.

Definition 3.5 (induced map) *Consider two complexes K, L and a simplicial map $f : K \to L$. We have a way to transform f into a map $f_\sharp : C_p(K) \to C_p(L)$, in this way*

$$c = \sum_{i=1}^{n} a_i \sigma_i \ \mapsto \ \sum_{i=1}^{??} a_i \tau_i$$

Where $\tau_i = f(\sigma_i)$ if it has dimension p or $\tau_i = 0$ otherwise. $f_\sharp$ is the induced map of f.