UNIVERSITY OF CALIFORNIA, BERKELEY

Lecture 18 - CS:189

Oscar Ortega

July 16, 2021

1 NEURONS

- CPUs: largely sequential, nanosecond gates, fragile if gate fails superior for arithmetic, logical rules, perfect key-based memories
- Brains: Very parallel, millisecond neurons, fault-tolerant. superior for vision, speech, associative memory. Noticing connections between things.
- Neuron: A cell in brain/nervous system for thinking and for communication
- Action Potentials: also known as spike, an electro-chemical impulse fired by a neuron to communicate w/other neurons as well as muscles and other glands
- Axon: the limbs along which the action potential propagates. "Output of the neuron".
- we know the action potential is caused by different action potentials moving in and out of the different parts of the axon.
- Dendrite: Smaller limbs by which neuron receives info; "input".
- Synapse: The connection from one neuron's axon to another neuron's dendrite.
- Neurotransmitter: Chemical released by axon terminal to stimulate dendrite

You have about 10^{11}, each with about 10^4 synapses.

1.1 ANALAOGIES

- output of unit $\iff$ firing rate of neuron
- Weight of connection $\iff$ synapse strength
- Positive weight $\iff$ excitatory neurotransmitter (e.g glutamine)
- Negative weight $\iff$ inhibitory neurotransmitter (e.g GABA, glycine)
- Linear combo of inputs $\iff$ summation of inputs
- Logistic/sigmoid fn $\iff$ there is a firing rate saturation
- Weight change/learning $\iff$ **synaptic plasticity** Hebb's rule (1949): "Cells that fire together, wire together"

We are pretty sure that brains don't perform any sort of back propagation.

2 MODULARITY OF THE BRAIN

- Frontal Lobe: composed of the frontal love and cerebellum. responsible for thinking, planning, problem solving, behavioral control, decision making
- brain stem: regulates body temperature, heart rate, swallowing, breathing
- temporal lobe: memory, understanding language, facial recognition, hearing, vision, speech, emotion
- occipital lobe: vision, visual processing, colour identification
- Parietal lobe: perception, object classification, spelling, knowledge of numbers, visuospacial processing.
- cerebellum: gross and fine motor skills, hand eye coordination, balance.

3 NEURAL NET VARIATIONS

3.1 REGRESSION

Usually linear output units - omit sigmoid fn. The derivative would be simpler.

3.2 Classification

What if we are dealing with a multi-class problem? in this case, we would go ahead and use the softmax function.
Let $t = Wh : t \in \mathbb{R}^k$ of a linear combination of the final layer.

$$\text{softmax} z_j(t) = \frac{e^{t_j}}{\sum_{i=1}^{k} e^{t_i}}$$

, we can think of this as counting the activation of the given component and normalizing. Corresponds to computing a probability for being in z_j

$$\forall j : \frac{\partial z_j}{\partial} t_j = z_j(1 - z_j) \tag{3.1}$$

$$\frac{\partial z_j}{\partial t_i} = -z_i z_j, j \neq i \tag{3.2}$$

$$\nabla_h z_j = z_j^T (W_j - W^T z) \tag{3.3}$$

3.3 Unit Saturation

Problem: When unit output s is close to 0 or 1 for most training point, $s' = s(1-s) \approx 0$, so gradient descent changes s very slowly. Unit is "stuck". Slowly training bad local minima.
To mitigate this we can do the following:

1. Initial weight of edge into unit with fan in η (where η is equal to the number of incoming edges: Random with mean zero, std. dev. $\frac{1}{\sqrt{\eta}}$
2. set target values to 0.15 0.85 instead of 0 1
3. modify backprop to add small constant (typically 0.1) to s'. Still choosing direction but not the steepest direction.
4. change the loss function, use the cross-entropy loss function. Version for sigmoid and for the soft-max
5. Replace sigmoids with ReLUs: **rectified Linear Units ramp fn**: aka hinge fn $r(\gamma) = \max\{0, \gamma\}$ $r'(\gamma) = \mathbb{1}(\gamma \geq 0)$, can cause the exploding gradients problem.

Recall that z and y corresponds to the predictions and the true values respectively.

in the two-class case:

$$L(z,y) = -\sum_i (y_i \ln(z_i) + (1-y_i)\ln(1-z_i)) \tag{3.4}$$

$$\frac{\partial L}{\partial z_j} = \frac{z_j - y_j}{z_j(1-z_j)} \tag{3.5}$$

$$\nabla_{w_j} L = \frac{\partial L}{\partial z_j} z_j(1-z_j)h \tag{3.6}$$

$$= (z_j - y_j)h \tag{3.7}$$

$$\nabla_h L = \sum_{j=1}^{k} \frac{\partial L}{\partial z_j} z_j(1-z_j)W_j = \sum_j (z_j - y_j)W_j = W^T(z-y) \tag{3.8}$$

For k-class softmax output, cross-entropy is the following:

$$L(z,y) = -\sum_{i=1}^{k} y_i \ln(z_i)$$