Artificial Intelligence Tutorial 6 - Answers

1. The nodes in a multi-layer neural network often contain sigmoid units, which perform the following calculation for a given weighted sum S:

$$\sigma(S) = \frac{1}{1+e^{-S}}$$

1a) Calculate the output from a sigmoid unit which takes the set {0.1, 0.8, 0.8, 0.3} as input.

1b) Is this unit "firing"?

> 1a) The weighted sum coming into the node is 0.1 + 0.8 + 0.8 + 0.3 = 2. e^{-2} is 0.135, so $1/(1+e^{-2})$ is 1/(1+0.135) = 0.88.
>
> 1b) This output is nearer to 1 than 0, so it can be thought of as firing.

Suppose we have a multi-layer network with two output nodes, and the target output for example E from output unit 1 is 1 and for output unit 2 is 0, yet the observed value produced for E was 0.3 for output unit 1 and 0.7 for output unit 2. The error term for an output unit O_k is calculated as:

$$\delta_{O_k} = o_k(E)(1 - o_k(E))(t_k(E) - o_k(E))$$

1c) What is the error term for O_1 and O_2?

> 1c) Remembering that $o_k(E)$ is the observed output from output unit k for training example E, and $t_k(E)$ is the target output from output unit k for E, the calculation is fairly straightforward:
>
> $d_{o1} = o_1(E)(1-o_1(E))(t_1(E)-o_1(E))$ = 0.3*(1-0.3)(1-0.3) = 0.3*0.7*0.7 = 0.147
>
> $d_{o2} = o_2(E)(1-o_2(E))(t_2(E)-o_2(E))$ = 0.7*(1-0.7)(0-0.7) = 0.7*0.3*-0.7 = -0.147

Suppose, in the ANN discussed in (b), the weight from hidden node H_1 to O_1 was 0.25 and from H_1 to O_2 it was 0.4. Suppose further that the output from H_1 for example E was 0.9.

1d) Use the following formula to calculate the error term of H_1 with respect to E:

$$\delta_{H_k} = h_k(E)(1 - h_k(E)) \sum_{i \in outputs} w_{ki}\delta_{O_i}$$

> 1d) To use this formula, we have to remember what each of the symbols represents. Firstly, in our case, k is 1, because we're dealing with the first hidden node (counting from top to bottom). Also, $h_k(E)$ is the observed output from hidden unit k when training example E is propagated through the network. Finally, w_{ki} is the weight between hidden unit k and output unit i and d_{oi} is the error term for output unit i, as calculated in part (b). The big sigma sign means that we add up all the values calculated for the different i's, where i ranges over the output units. We will deal first with the

summation. There are two output nodes, and the error terms for them were 0.147 and –0.147 respectively, with the weights between hidden unit 1 and them being 0.25 and 0.4 respectively (as specified in the question). Hence, the summation is:

0.25*0.147 + 0.4*(-0.147) = -0.02205

Next, we can calculate that $h_1(E)(1-h_1(E))$ is 0.9*(1-0.9) = 0.9 * 0.1 = 0.09. Finally, we can put the two halves of the calculation together to give:

d_{hk} = -0.2205*0.09 = -0.0019845

Suppose in the last epoch of training using example E, the weight change for h_1 was 0.004. Suppose further that we are using a learning rate of 0.1 and a momentum of 0.2 in training the network.

1e) What will be the change for the weight between h_1 and o_1 using example E to train it in the current epoch?

1e) The weight change is calculated by multiplying the error term for the output unit by the learning rate (given as 0.1), then multiplying the answer by the output from the hidden unit for example E. We have calculated the error rate in part (b) to be 0.147, and we also know that the output from H1 for E was 0.7. Therefore, the weight change is:

Δ_{11} = 0.1 * 0.147 * 0.9 = 0.01323

Momentum is a heuristic used to get over local minimas, and it works by adding on a fraction of the previous weight change during the current weight change. The fraction is called the momentum, and in our case it was 0.2. Hence, to alter this weight change using momentum, we need to add on 0.2 times the previous weight change, which we were told was 0.004. Hence, our final weight change is:

Δ_{11} = 0.01323 + 0.2*0.004 = 0.01403

2. Suppose you know one thing in life:

(i) all people who smoke are stupid

You've also observed that

(ii) people who smoke and are smelly are ugly.

2a) Use the absorption rule to induce a hypothesis about stupid, smelly, ugly people (i.e., a rule not involving smoking).

2a) The absorption rule of inductive inference can be portrayed like this:

$$\frac{q \leftarrow A \;\&\; p \leftarrow A,B}{q \leftarrow A \;\&\; p \leftarrow q,B}$$

Where the commas represent conjunction between (sets of) literals, p and q are single literals and A and B are conjunctions of literals. Note that A and B could also just be single literals. We need to choose which literals (smokes(X), stupid(X), smelly(X) and ugly(X)) to put into the above rule in place of p, q, A and B. We know that the clause on the bottom right is the new one induced by this rule and that we want to get rid of the smoking literal. In the bottom right literal, A does not appear, so we should make A be the single literal smokes(X). This suggests that we make q the literal stupid, so that we start with the rule q ← A being: stupid(X) ← smokes(X), which states that anyone who smokes is stupid. We also want to write down the second rule: people who smoke and are smelly are ugly. This suggests that we make literal p be ugly(X) and B be the single literal smelly(X). This makes the top line of the inference look like this:

stupid(X) ← smokes(X) & ugly(X) ← smokes(X), smelly(X)

From this, we can induce the bottom line of the rule as:

stupid(X) ← smokes(X) & ugly(X) ← stupid(X), smelly(X)

Of course, the clause on the right is the new one that we have inductively inferred. This states that all stupid, smelly people are ugly.

2b) Translate your induced hypothesis into CNF (you should be getting good at this by now) in propositional logic. Also translate sentence (i). Use a resolution step to show that your induced hypothesis, when taken to be true alongside sentence (i) can be used to explain sentence (ii).

2b) We are allowed to work in propositional logic, so we don't need to worry about variables, etc. We will use the propositions "A person is stupid", written `stupid`, "A person is ugly", written `ugly`, "A person is smelly", written `smelly` and "A person smokes", written `smokes`. Our original sentence (i) stated that:

(i) `smoke → stupid`

and our hypothesis, that stupid, smelly people are ugly can be written as:

(H) `stupid ∧ smelly → ugly`

Re-writing these into CNF, we get:

(i) `¬smoke v stupid`

(H) `¬(stupid ∧ smelly) v ugly`
which needs to be further re-written as:
`¬stupid v ¬smelly v ugly`

We see that stupid appears in (i) and ¬stupid appears in (H), so we can now resolve the stupid proposition away as follows:

(i) `¬smoke v stupid`   (H) `¬stupid v ¬smelly v ugly`

`¬smoke v ¬smelly v ugly`

This needs a little re-writing into: `¬(smoke v ¬smelly) v ugly`
and finally into: `(smoke ∧ smelly) → ugly`

Hence, we see that given (i) and (H) as true, we can deduce using resolution that people who smoke and are smelly are ugly, which is exactly what we observed and wrote down in sentence (ii).

In English, we have done the following: we stated that we knew that people who smoke are stupid. We didn't know why, but we've noticed that smokers who are smelly are also ugly. This could possibly be because stupid, smelly people are ugly. If this were true, then because smokers are known to be stupid, then smelly smokers are stupid smelly people, and, by our conjecture are ugly – so the hypothesis explains our observation.