Confidence Measures

Confidence

- Suppose that we must design a neural network that resides in a fighter plane, examining the radar return, signature of distant objects.
- The purpose of our network is to help the pilot judge whether or not a blip on the screen is an aircraft.
- She would be dismayed if the network simply reported that a blip that just appeared had activated the network's output neuron to a 73-percent activation level! It would be far more informative for the network to report that there is a 92-percent chance that the blip is from an aircraft.

Confidence

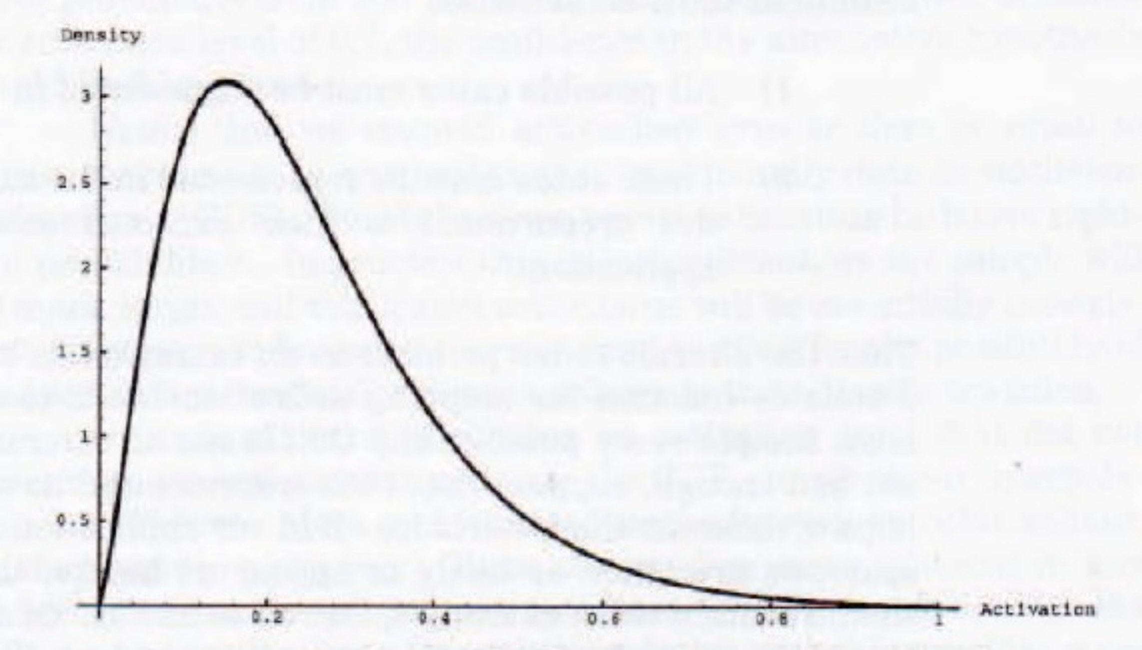
- How do we go about converting that 73-percent activation to 92-percent probability?
- Two ways: The first is based on the standard statistical concept of ***hypothesis testing*** and will deal primarily with the negative task of rebutting decisions. The second is based on ***Bayes inference***.

Testing Individual Hypotheses

- Two hypotheses: One is called ***null*** hypothesis (blip IS NOT an aircraft); the other one is called ***alternative*** hypothesis (blip IS an aircraft)
- *Type-I error*: We may falsely reject the null hypothesis by asserting that the blip is an aircraft when, in fact, it is not…
- *Type-II error*: Falsely accepting the null hypothesis, concluding that the blip is not an aircraft when, in fact it is.

Null Hypothesis

- The confidence level in the alternative hypothesis is 1 – the probability of making a Type-I error. The probability that a non-aircraft could have been responsible for a level this high (or higher) is equal to the area under this curve to the right of the activation level.
- As can be seen, activation levels approaching 1 have vanishing area to their right. Hence have alternative-hypothesis confidence approaching 1.



Typical activations under null hypothesis

Computing Confidence

- For example, suppose that our (ridiculously small) collection of nonaircraft blip activation levels consists of 0.1, 0.2, 0.2, 0.4, and 0.7, for a total of five samples.
- Then we see that two out of five (40 percent) of our sample achieved an activation level of at least 0.4, four out of five (80 percent) were at 0.2 or more, et cetera. This percentage gives the probability of a ***type I error***.
- The confidence level in the alternative hypothesis is computed by subtracting the type I error probability from 100 percent. Thus, if our application achieved an activation level of 0.7, the confidence in the alternative hypothesis would be 80 percent.

Confidence in the Null Hypothesis

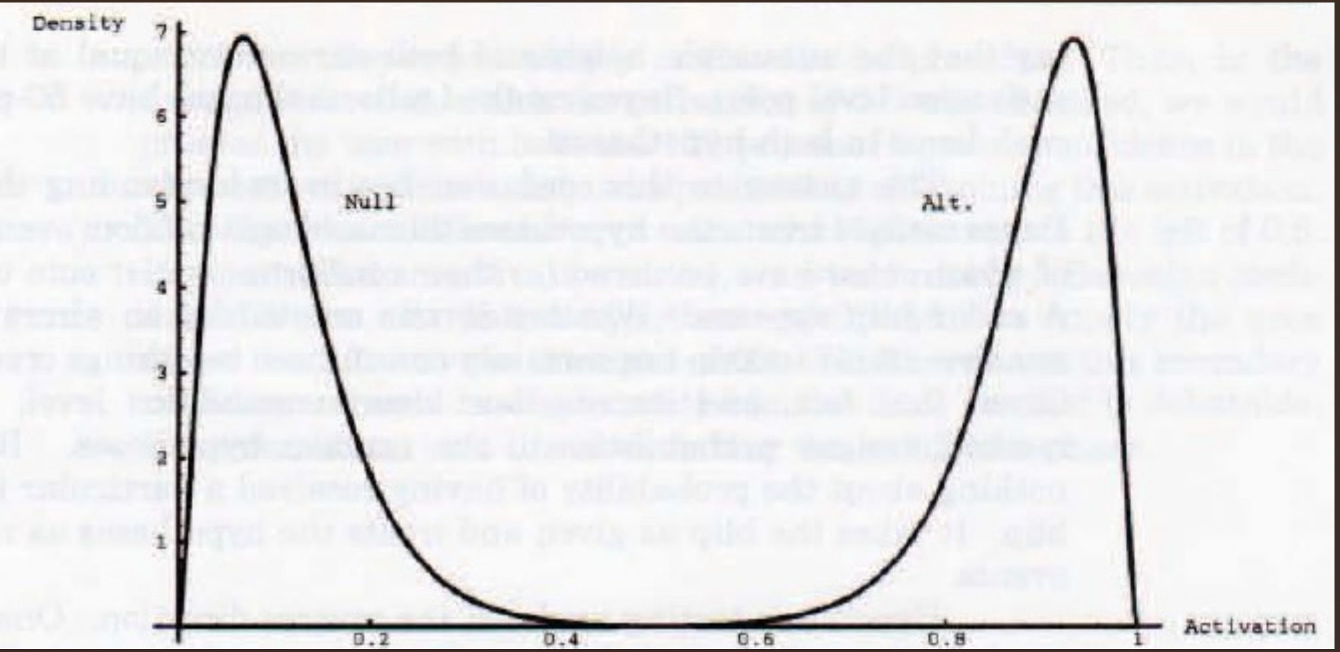
- Suppose our collection of aircraft consists of activations of 0.5,0.7,0.7,0.8, and 0.9. If an activation of 0.7 were attained, 3 out of the 5 samples are less than or equal to it, so it then maps to a confidence of 40 percent (1.0 - 0.6) in the null (non-aircraft) hypothesis. Activations less than or equal to 0.5 map to a confidence of 80 percent in the null hypothesis.

Multiple Classes

- The previous discussion focused on the case of a simple decision: either the null hypothesis is true, or the alternative hypothesis is true.
- Confidence computation can be generalized to the case of multiple alternative hypotheses. A radar blip may be a non-aircraft, fighter, or passenger plane. Things get so much more difficult, though, that it is rarely worth the effort. The Bayes method for probabilistic neural networks is usually better if there is more than one alternative.

Hypothesis Testing versus Bayes' Method

- If the network achieves an activation level of 0.1, the two competing confidence measures will be numerically different, but identical in meaning. The confidence in rejecting the null hypothesis is about 50 percent (just eyeballing the left curve), which tells us nothing useful. However. the confidence in rejecting the alternative hypothesis is essentially 100 percent, because virtually all of the area under the right curve lies to the right of 0.1.
- The confidence in the alternative hypothesis as computed with the Bayes method, which is the height of the right curve at 0.1 divided by the sum of the heights of the left and right curves at that point, is clearly 0. Hence the Bayes confidence in the null hypothesis is 100 percent. Similarly, there would be no problem for activations near 1.



Hypothesis Testing versus Bayes' Method

- Now let us examine the case of an activation level of 0.3. The confidence in rejecting the alternative hypothesis, and the Bayes confidence in the null hypothesis, will still both be 100 percent. But now the confidence in rejecting the null hypothesis is greater than 95 percent, for less than 5 percent of the area under the left curve lies to the right of 0.3.
- Bayes' theorem tells us that the confidence in the null hypothesis is 100 percent, while hypothesis testing tells us that we can have better than a 95-percent confidence in *rejecting* the null hypothesis! What's up with this?

Hypothesis Testing versus Bayes' Method

- If our network achieves an activation level of 0.5, things get really confusing. Hypothesis testing will give us virtually l00-percent confidence in rejecting *both* the null and alternative hypotheses.
- Supposing that the minuscule heights of both curves are equal at the 0.5 activation level point, Bayes' method tells us that we have 50-percent confidence in both hypotheses!

Bayes method

- The answer to this confusion lies in understanding that the Bayes method treats the hypotheses themselves as random events, one of which *must* have occurred (so their confidences *must* sum to one). A radar blip appears. Whether it was created by an aircraft or a non-aircraft is random, but certainly one of those two things created it.
- Given that fact, and its resultant neuron activation level, Bayes' method assigns probabilities to the random hypotheses. It cares nothing about the probability of having received a particular form of blip. It takes the blip as given and treats the hypotheses as random events.

Hypothesis testing method

- Hypothesis testing works in the reverse direction. One of the two hypotheses is given to be true. It is not a random event. It is the blip that is the random event.
- The question answered concerns the probability of getting the attained neuron activation level under the assumption of a particular hypothesis. *This probability is not addressed in Bayes' method.*

Finally

- So, in the final example above, in which a 50-percent activation level was attained. Bayes' method ignores the fact that this is a practical impossibility. It just says that there is a 50-50 chance that either hypothesis could be true. Hypothesis testing, on the other hand, recognizes the rareness of that event and tells us to reject both hypotheses.

Which method to choose?

- Use both to generate two confidences. The primary confidence would be the Bayes. However, the probability of being wrong under the hypothesis *that has maximum Bayes probability* could also be provided.
- Say, an activation of 0.6 was obtained. We would present the user with both the 100-percent Bayesian confidence in the alternative, and the nearly zero probability of attaining this activation. This number would be the area under the right curve to the left of 0.6.