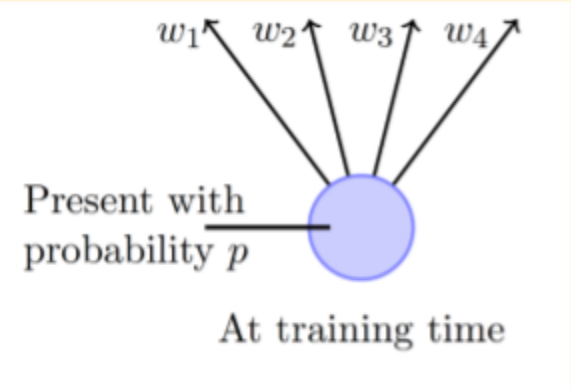
PadhAI: Batch Normalization and Dropout

One Fourth Labs

Using dropout at test time

How do you use the networks at test time?

1. At the end of training, we are left with the outputs of the various networks, that we combine with an ensemble method such as average, voting etc.
2. In theory, combining 2^N such outputs would prove to be very computationally expensive.
3. There is a simple strategy to overcome this.
4. Consider any particular neuron from a dropped-network during training.

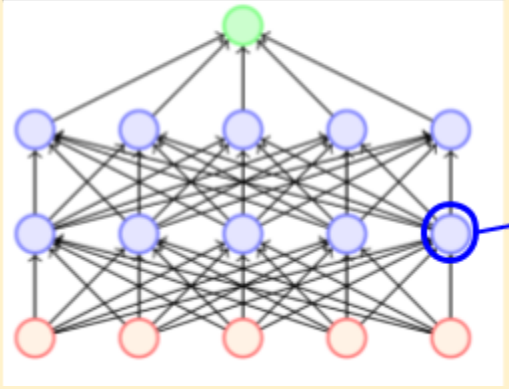


In any of the dropped-out networks, the neuron has p probability of existing.

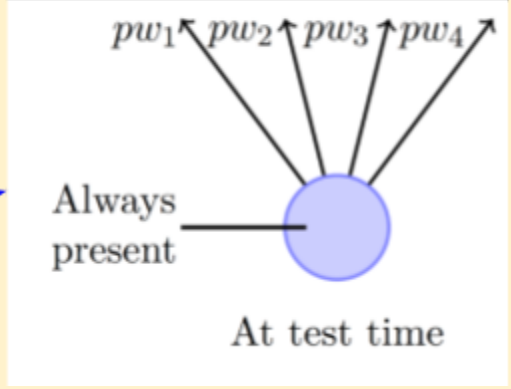
So far, we have assumed p = 0.5 or 50%
However p can take other values

5. Now, during test time, instead of using the ensemble, we use the large mega-network.
6. Instead of each neuron being present with probability p, we assume that all neurons are present at all times.
7. However, we multiply the output of the neurons with the probability value p.

Test using the mega-network



Multiply outputs with p



8. We scale the output of the neuron by p to show that it is only p% reliable.
9. So in our example of p = 0.5, it's like saying that a particular neuron is only present in 50% of the networks, thus we approximate it by multiplying its outputs by 0.5