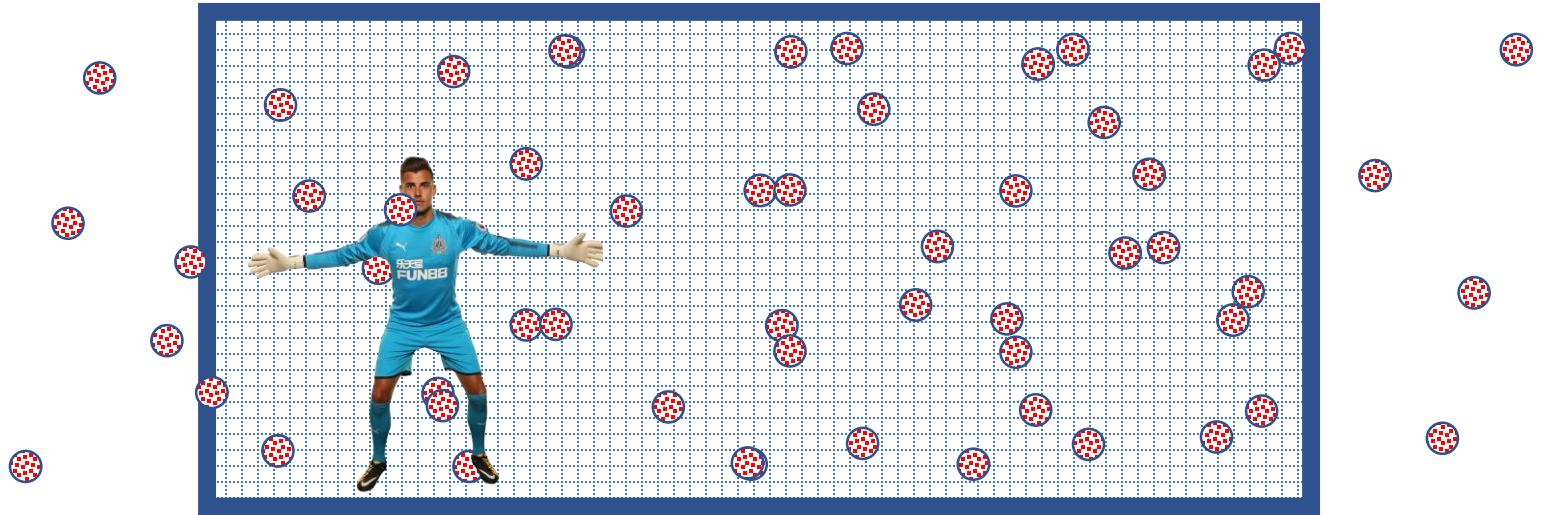


# Gradient Backpropagation Tuning

Dai Bui

# Batch Normalization

- Remember that we train in batch
  - What happens if the mean and variance of data are shifted significantly after each batch?
- Imagine training a goalkeeper
  - Mean of training shots are in the left, train more with left side
  - Mean of training shots are in the right, train more with right side
  - Both cases, the goalkeeper needs to switch between different skills and states, maybe he would forget his previous skills when training in the new batch
  - What we want to do?
    - Shift the training shots so that the mean is in the middle



# Batch Normalization

- For a layer with  $d$  dimensional vector  $x = (x_1, x_2, \dots, x_d)$ , we normalize each dimension separately as follows:

$$\hat{x}_k = \frac{x_k - E[x_k]}{\sqrt{\text{Var}[x_k]}}$$

- For a mini-batch  $B = \{x^1, x^2, \dots, x^m\}$  of  $m$  samples, we do the following normalization to transform into  $\{$

$$\mu_{Bi} = \frac{1}{m} \sum_{j=1}^m x_i^j$$

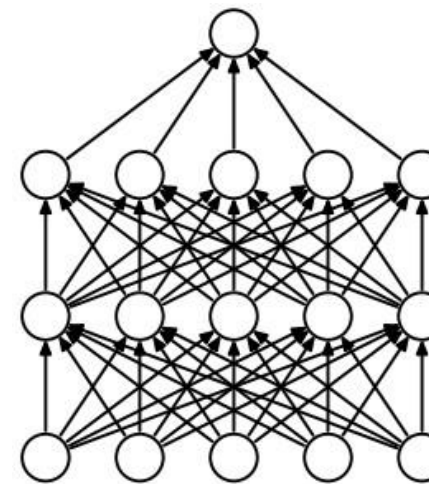
$$\hat{x}_i^j = \frac{x_i^j - \mu_{Bi}}{\sqrt{\sigma_{Bi}^2 - \epsilon}}$$

$$\sigma_{Bi}^2 = \frac{1}{m} \sum_{j=1}^m (x_i^j - \mu_{Bi})^2$$

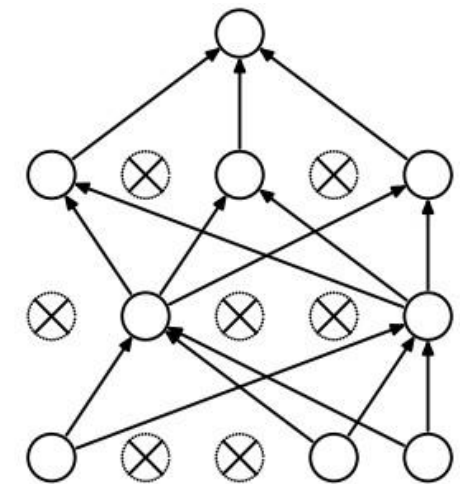
$$y^j = \gamma \hat{x}_i^j + \beta$$

# Dropout

- What happens if a company depends too much on a small number of employees?
- What happens if a project group depends too much on a single member?
- How what happens if those employees get sick?
- How do we prepare for such an event?



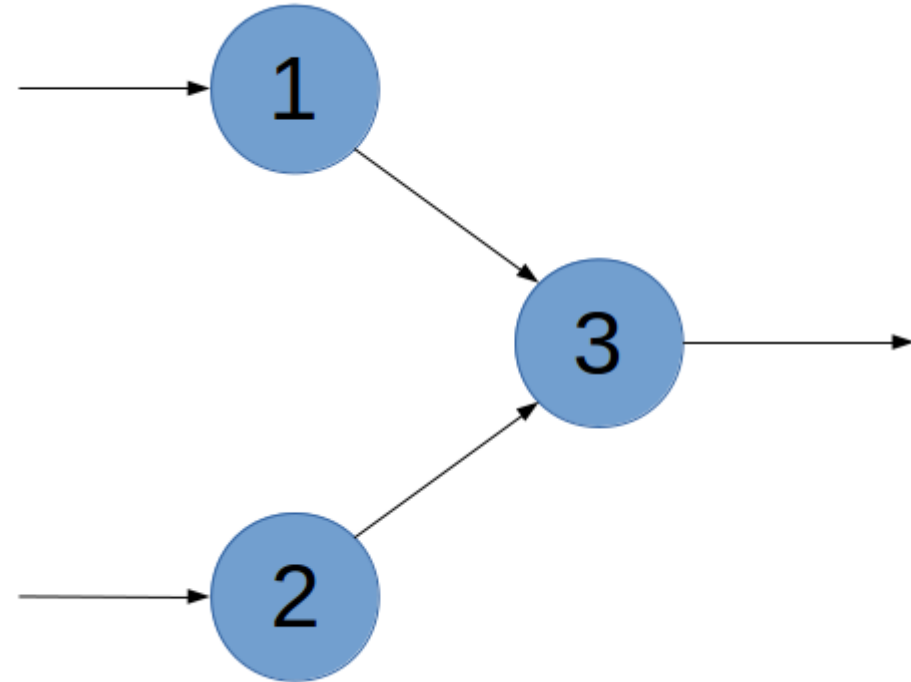
(a) Standard Neural Net



(b) After applying dropout.

# Dropout example

- Neuron 1 produces good output that gives 80% correct prediction
- Neuron 2 produces random output
- Then Neuron 3 would do
  - $out_3 = 1 * out_1 + 0 * out_2$
- If we want to train Neuron 2, we need to drop Neuron 1
  - Neuron 2 would stop produce random outputs



# Practice: Batch Normalization and Dropout