Deep Learning

Vazgen Mikayelyan

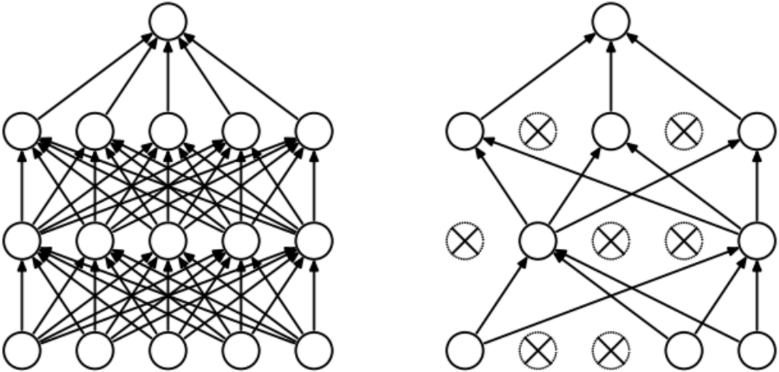
YSU, Krisp

October 14, 2020

Outline

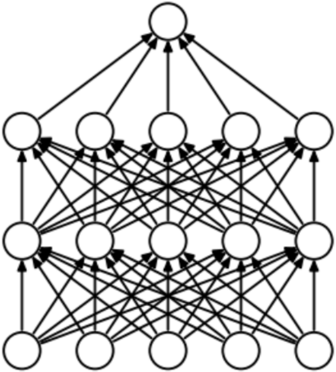
Dropout

Dropout

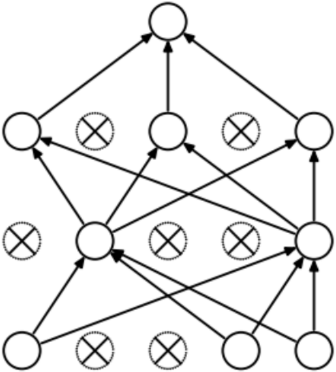


(a) Standard Neural Net (b) After applying dropout.

Dropout



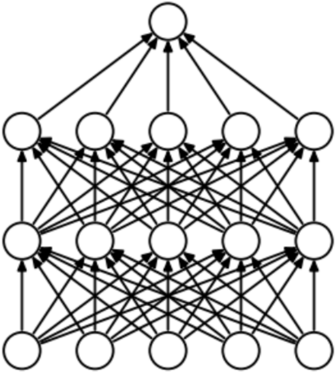
(a) Standard Neural Net



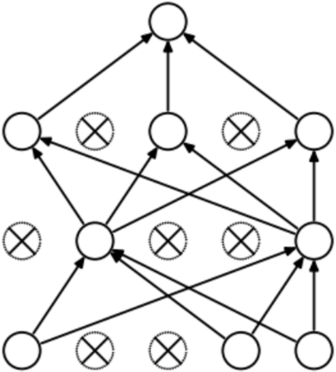
(b) After applying dropout.

What to do during the inference?

Dropout



(a) Standard Neural Net



(b) After applying dropout.

What to do during the inference?

Answer: Scale units by $\dfrac{1}{1 - rate}$ during the training and set rate=1 during the inference.

Outline

Simple Moving Average

Definition 1

Simple moving average of the given data is the arithmetic mean of the previous k data.

Definition 1

Simple moving average of the given data is the arithmetic mean of the previous k data.

If you have the data $x_1, x_2, \ldots$, then its simple moving average will be the following

$$\mu_n = \frac{x_{n-k+1} + \ldots + x_n}{k}, n = k, k+1, \ldots.$$

Cumulative Moving Average

Definition 2

Cumulative moving average of the given data is the arithmetic mean of the all previous data up to the current time.

Cumulative Moving Average

Definition 2

Cumulative moving average of the given data is the arithmetic mean of the all previous data up to the current time.

If you have the data $x_1, x_2, \ldots$, then its cumulative moving average will be the following

$$\mu_n = \frac{x_1 + x_2 + \ldots + x_n}{n}, n = 1, 2, \ldots .$$

Definition 2

Cumulative moving average of the given data is the arithmetic mean of the all previous data up to the current time.

If you have the data $x_1, x_2, \ldots$, then its cumulative moving average will be the following

$$\mu_n = \frac{x_1 + x_2 + \ldots + x_n}{n}, n = 1, 2, \ldots .$$

Note that

$$\mu_n = \frac{(x_1 + x_2 + \ldots + x_{n-1}) + x_n}{n} = \frac{(n-1)\,\mu_{n-1} + x_n}{n}$$

$$= \left(1 - \frac{1}{n}\right)\mu_{n-1} + \frac{1}{n} x_n.$$

Exponential Moving Average

If you have the data $x_1, x_2, \ldots$, then its exponential moving average will be the following

$$\mu_1 = x_1,$$

$$\mu_n = \alpha \mu_{n-1} + (1-\alpha)\, x_n, \;\; n \geq 2$$

Exponential Moving Average

If you have the data $x_1, x_2, \ldots$, then its exponential moving average will be the following

$$\mu_1 = x_1,$$

$$\mu_n = \alpha\mu_{n-1} + (1-\alpha)\,x_n,\;\; n \geq 2$$

Note that

$$\mu_n = \alpha\mu_{n-1} + (1-\alpha)\,x_n = \alpha\left(\alpha\mu_{n-2} + (1-\alpha)\,x_{n-1}\right) + (1-\alpha)\,x_n$$

Exponential Moving Average

If you have the data $x_1, x_2, \ldots$, then its exponential moving average will be the following

$$\mu_1 = x_1,$$

$$\mu_n = \alpha \mu_{n-1} + (1-\alpha) x_n, \;\; n \geq 2$$

Note that

$$\mu_n = \alpha \mu_{n-1} + (1-\alpha) x_n = \alpha \left(\alpha \mu_{n-2} + (1-\alpha) x_{n-1} \right) + (1-\alpha) x_n$$

$$= \alpha^2 \mu_{n-2} + (1-\alpha) \alpha x_{n-1} + (1-\alpha) x_n$$

Exponential Moving Average

If you have the data $x_1, x_2, \ldots$, then its exponential moving average will be the following

$$\mu_1 = x_1,$$

$$\mu_n = \alpha \mu_{n-1} + (1-\alpha)\, x_n, \;\; n \geq 2$$

Note that

$$\mu_n = \alpha \mu_{n-1} + (1-\alpha)\, x_n = \alpha\left(\alpha \mu_{n-2} + (1-\alpha)\, x_{n-1}\right) + (1-\alpha)\, x_n$$

$$= \alpha^2 \mu_{n-2} + (1-\alpha)\, \alpha x_{n-1} + (1-\alpha)\, x_n$$

$$= \alpha^3 \mu_{n-3} + (1-\alpha)\, \alpha^2 x_{n-2} + (1-\alpha)\, \alpha x_{n-1} + (1-\alpha)\, x_n$$

Exponential Moving Average

If you have the data $x_1, x_2, \ldots$, then its exponential moving average will be the following

$$\mu_1 = x_1,$$

$$\mu_n = \alpha\mu_{n-1} + (1-\alpha)\,x_n,\ \ n \geq 2$$

Note that

$$\mu_n = \alpha\mu_{n-1} + (1-\alpha)\,x_n = \alpha\left(\alpha\mu_{n-2} + (1-\alpha)\,x_{n-1}\right) + (1-\alpha)\,x_n$$

$$= \alpha^2\mu_{n-2} + (1-\alpha)\,\alpha x_{n-1} + (1-\alpha)\,x_n$$

$$= \alpha^3\mu_{n-3} + (1-\alpha)\,\alpha^2 x_{n-2} + (1-\alpha)\,\alpha x_{n-1} + (1-\alpha)\,x_n$$

$$= \alpha^{n-1}\mu_1 + (1-\alpha)\,\alpha^{n-2}x_2 + \ldots + (1-\alpha)\,\alpha x_{n-1} + (1-\alpha)\,x_n.$$

Exponential Moving Average

If you have the data $x_1, x_2, \ldots$, then its exponential moving average will be the following

$$\mu_1 = x_1,$$

$$\mu_n = \alpha\mu_{n-1} + (1-\alpha)\,x_n, \;\; n \geq 2$$

Note that

$$\mu_n = \alpha\mu_{n-1} + (1-\alpha)\,x_n = \alpha\left(\alpha\mu_{n-2} + (1-\alpha)\,x_{n-1}\right) + (1-\alpha)\,x_n$$

$$= \alpha^2\mu_{n-2} + (1-\alpha)\,\alpha x_{n-1} + (1-\alpha)\,x_n$$

$$= \alpha^3\mu_{n-3} + (1-\alpha)\,\alpha^2 x_{n-2} + (1-\alpha)\,\alpha x_{n-1} + (1-\alpha)\,x_n$$

$$= \alpha^{n-1}\mu_1 + (1-\alpha)\,\alpha^{n-2}x_2 + \ldots + (1-\alpha)\,\alpha x_{n-1} + (1-\alpha)\,x_n.$$

It easy to see that sum of the coefficients is equal to 1.

Outline

Batch Normalization

- Problem:
  - The distribution of each layer's input changes during training.

Batch Normalization

- Problem:
  - The distribution of each layer's input changes during training.
- Solution:
  - Fix the distribution of inputs into subnetwork.

Batch Normalization

- Problem:
  - The distribution of each layer's input changes during training.
- Solution:
  - Fix the distribution of inputs into subnetwork.
- Effects:
  - Improve accuracy.
  - Faster learning.
  - Availability of high learning rates.

Batch Normalization

Input: Values of x over a mini-batch: $\mathcal{B} = \{x_{1...m}\}$;
Parameters to be learned: γ, β

Output: $\{y_i = \mathrm{BN}_{\gamma,\beta}(x_i)\}$

$$\mu_{\mathcal{B}} \leftarrow \frac{1}{m}\sum_{i=1}^{m} x_i \qquad \text{// mini-batch mean}$$

$$\sigma_{\mathcal{B}}^2 \leftarrow \frac{1}{m}\sum_{i=1}^{m} (x_i - \mu_{\mathcal{B}})^2 \qquad \text{// mini-batch variance}$$

$$\widehat{x}_i \leftarrow \frac{x_i - \mu_{\mathcal{B}}}{\sqrt{\sigma_{\mathcal{B}}^2 + \epsilon}} \qquad \text{// normalize}$$

$$y_i \leftarrow \gamma \widehat{x}_i + \beta \equiv \mathrm{BN}_{\gamma,\beta}(x_i) \qquad \text{// scale and shift}$$