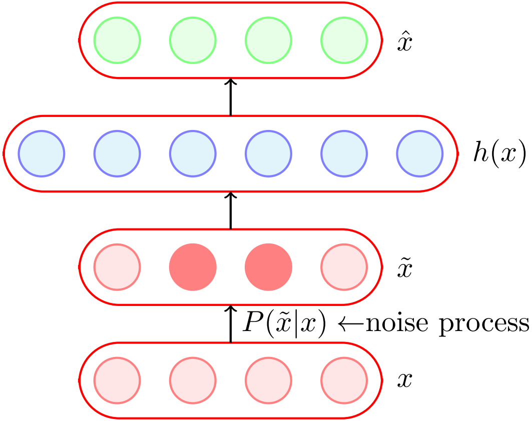
Module 8.7 : Adding Noise to the inputs

Other forms of regularization

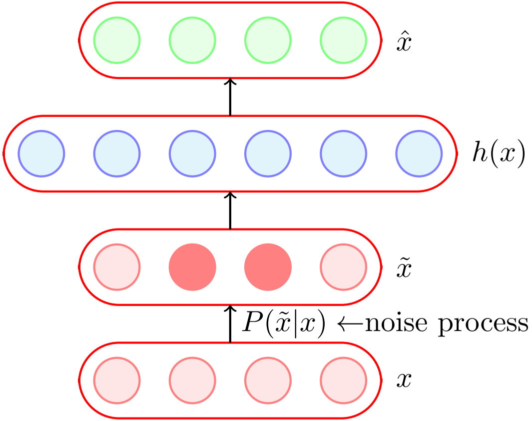
- l_2 regularization
- Dataset augmentation
- Parameter Sharing and tying
- Adding Noise to the inputs
- Adding Noise to the outputs
- Early stopping
- Ensemble methods
- Dropout

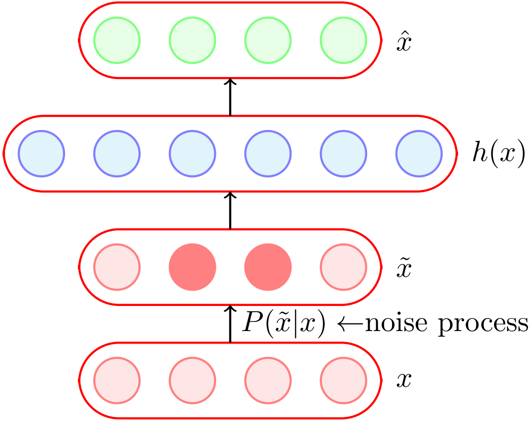
Other forms of regularization

- l_2 regularization
- Dataset augmentation
- Parameter Sharing and tying
- Adding Noise to the inputs
- Adding Noise to the outputs
- Early stopping
- Ensemble methods
- Dropout

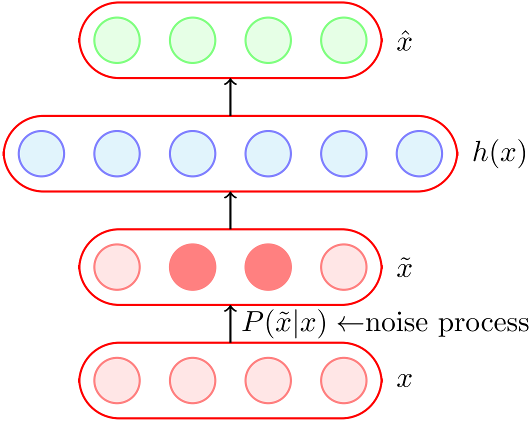

$\hat{x}$
$h(x)$
$\tilde{x}$
$P(\tilde{x}|x)$ ←noise process
x

- We saw this in Autoencoder

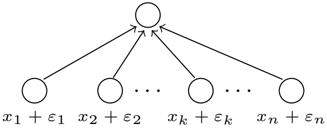




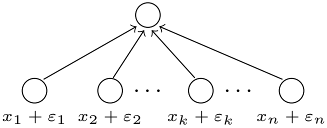
- We saw this in Autoencoder
- We can show that for a simple input output neural network, adding Gaussian noise to the input is equivalent to weight decay (L_2 regularisation)



- We saw this in Autoencoder
- We can show that for a simple input output neural network, adding Gaussian noise to the input is equivalent to weight decay (L_2 regularisation)
- Can be viewed as data augmentation

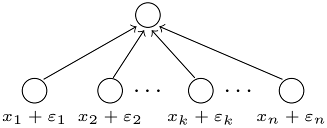


$$\varepsilon \sim \mathcal{N}(0, \sigma^2)$$



$$\varepsilon \sim \mathcal{N}(0, \sigma^2)$$

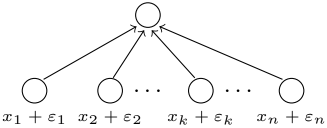
$$\widetilde{x}_i = x_i + \varepsilon_i$$



$$\varepsilon \sim \mathcal{N}(0, \sigma^2)$$

$$\widetilde{x}_i = x_i + \varepsilon_i$$

$$\widehat{y} = \sum_{i=1}^{n} w_i x_i$$

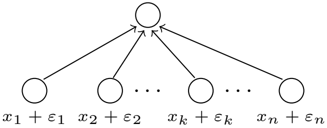


$$\varepsilon \sim \mathcal{N}(0, \sigma^2)$$

$$\widetilde{x}_i = x_i + \varepsilon_i$$

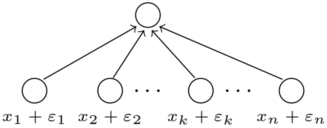
$$\widehat{y} = \sum_{i=1}^{n} w_i x_i$$

$$\widetilde{y} = \sum_{i=1}^{n} w_i \widetilde{x}_i$$



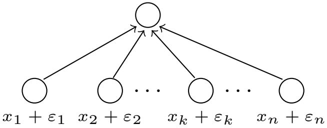
$$\varepsilon \sim \mathcal{N}(0, \sigma^2)$$

$$\begin{aligned}
\widetilde{x}_i &= x_i + \varepsilon_i \\
\widehat{y} &= \sum_{i=1}^{n} w_i x_i \\
\widetilde{y} &= \sum_{i=1}^{n} w_i \widetilde{x}_i \\
&= \sum_{i=1}^{n} w_i x_i + \sum_{i=1}^{n} w_i \varepsilon_i
\end{aligned}$$



$$\varepsilon \sim \mathcal{N}(0, \sigma^2)$$

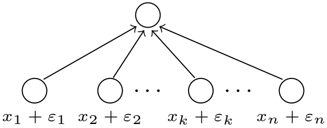
$$\begin{aligned}
\widetilde{x_i} &= x_i + \varepsilon_i \\
\widehat{y} &= \sum_{i=1}^{n} w_i x_i \\
\widetilde{y} &= \sum_{i=1}^{n} w_i \widetilde{x_i} \\
&= \sum_{i=1}^{n} w_i x_i + \sum_{i=1}^{n} w_i \varepsilon_i \\
&= \widehat{y} + \sum_{i=1}^{n} w_i \varepsilon_i
\end{aligned}$$



$$\varepsilon \sim \mathcal{N}(0, \sigma^2)$$

$$\begin{aligned}
\widetilde{x}_i &= x_i + \varepsilon_i \\
\widehat{y} &= \sum_{i=1}^{n} w_i x_i \\
\widetilde{y} &= \sum_{i=1}^{n} w_i \widetilde{x}_i \\
&= \sum_{i=1}^{n} w_i x_i + \sum_{i=1}^{n} w_i \varepsilon_i \\
&= \widehat{y} + \sum_{i=1}^{n} w_i \varepsilon_i
\end{aligned}$$

We are interested in $E[(\widetilde{y} - y)^2]$



$$\varepsilon \sim \mathcal{N}(0, \sigma^2)$$

$$\widetilde{x_i} = x_i + \varepsilon_i$$

$$\widehat{y} = \sum_{i=1}^{n} w_i x_i$$

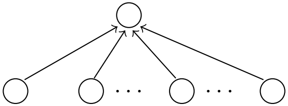
$$\widetilde{y} = \sum_{i=1}^{n} w_i \widetilde{x_i}$$

$$= \sum_{i=1}^{n} w_i x_i + \sum_{i=1}^{n} w_i \varepsilon_i$$

$$= \widehat{y} + \sum_{i=1}^{n} w_i \varepsilon_i$$

We are interested in $E[(\widetilde{y} - y)^2]$

$$E\left[(\widetilde{y} - y)^2\right] = E\left[\left(\widehat{y} + \sum_{i=1}^{n} w_i \varepsilon_i - y\right)^2\right]$$



$$\varepsilon \sim \mathcal{N}(0, \sigma^2)$$

$$\widetilde{x}_i = x_i + \varepsilon_i$$

$$\widehat{y} = \sum_{i=1}^{n} w_i x_i$$

$$\widetilde{y} = \sum_{i=1}^{n} w_i \widetilde{x}_i$$

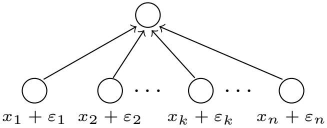
$$= \sum_{i=1}^{n} w_i x_i + \sum_{i=1}^{n} w_i \varepsilon_i$$

$$= \widehat{y} + \sum_{i=1}^{n} w_i \varepsilon_i$$

We are interested in $E[(\widetilde{y} - y)^2]$

$$E\left[(\widetilde{y} - y)^2\right] = E\left[\left(\widehat{y} + \sum_{i=1}^{n} w_i \varepsilon_i - y\right)^2\right]$$

$$= E\left[\left(\left(\widehat{y} - y\right) + \left(\sum_{i=1}^{n} w_i \varepsilon_i\right)\right)^2\right]$$



$$\varepsilon \sim \mathcal{N}(0, \sigma^2)$$

$$\widetilde{x}_i = x_i + \varepsilon_i$$

$$\widehat{y} = \sum_{i=1}^{n} w_i x_i$$

$$\widetilde{y} = \sum_{i=1}^{n} w_i \widetilde{x}_i$$

$$= \sum_{i=1}^{n} w_i x_i + \sum_{i=1}^{n} w_i \varepsilon_i$$

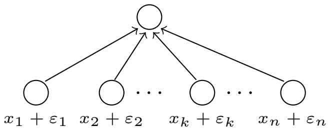
$$= \widehat{y} + \sum_{i=1}^{n} w_i \varepsilon_i$$

We are interested in $E[(\widetilde{y} - y)^2]$

$$E\left[(\widetilde{y} - y)^2\right] = E\left[\left(\widehat{y} + \sum_{i=1}^{n} w_i \varepsilon_i - y\right)^2\right]$$

$$= E\left[\left(\left(\widehat{y} - y\right) + \left(\sum_{i=1}^{n} w_i \varepsilon_i\right)\right)^2\right]$$

$$= E\left[(\widehat{y} - y)^2\right] + E\left[2(\widehat{y} - y)\sum_{i=1}^{n} w_i \varepsilon_i\right] + E\left[\left(\sum_{i=1}^{n} w_i \varepsilon_i\right)^2\right]$$



$$\varepsilon \sim \mathcal{N}(0, \sigma^2)$$

$$\widetilde{x_i} = x_i + \varepsilon_i$$

$$\widehat{y} = \sum_{i=1}^{n} w_i x_i$$

$$\widetilde{y} = \sum_{i=1}^{n} w_i \widetilde{x_i}$$

$$= \sum_{i=1}^{n} w_i x_i + \sum_{i=1}^{n} w_i \varepsilon_i$$

$$= \widehat{y} + \sum_{i=1}^{n} w_i \varepsilon_i$$

We are interested in $E[(\widetilde{y} - y)^2]$

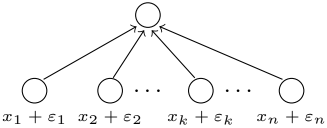
$$E\left[(\widetilde{y} - y)^2\right] = E\left[\left(\widehat{y} + \sum_{i=1}^{n} w_i \varepsilon_i - y\right)^2\right]$$

$$= E\left[\left(\left(\widehat{y} - y\right) + \left(\sum_{i=1}^{n} w_i \varepsilon_i\right)\right)^2\right]$$

$$= E\left[(\widehat{y} - y)^2\right] + E\left[2(\widehat{y} - y)\sum_{i=1}^{n} w_i \varepsilon_i\right] + E\left[\left(\sum_{i=1}^{n} w_i \varepsilon_i\right)^2\right]$$

$$= E\left[(\widehat{y} - y)^2\right] + 0 + E\left[\sum_{i=1}^{n} w_i^2 \varepsilon_i^2\right]$$

($\because$ ε_i is independent of ε_j and ε_i is independent of ($\widehat{y}$-y))



$$\varepsilon \sim \mathcal{N}(0, \sigma^2)$$

$$\widetilde{x}_i = x_i + \varepsilon_i$$

$$\widehat{y} = \sum_{i=1}^{n} w_i x_i$$

$$\widetilde{y} = \sum_{i=1}^{n} w_i \widetilde{x}_i$$

$$= \sum_{i=1}^{n} w_i x_i + \sum_{i=1}^{n} w_i \varepsilon_i$$

$$= \widehat{y} + \sum_{i=1}^{n} w_i \varepsilon_i$$

We are interested in $E[(\widetilde{y} - y)^2]$

$$E\left[(\widetilde{y} - y)^2\right] = E\left[\left(\widehat{y} + \sum_{i=1}^{n} w_i \varepsilon_i - y\right)^2\right]$$

$$= E\left[\left(\left(\widehat{y} - y\right) + \left(\sum_{i=1}^{n} w_i \varepsilon_i\right)\right)^2\right]$$

$$= E\left[(\widehat{y} - y)^2\right] + E\left[2(\widehat{y} - y)\sum_{i=1}^{n} w_i \varepsilon_i\right] + E\left[\left(\sum_{i=1}^{n} w_i \varepsilon_i\right)^2\right]$$

$$= E\left[(\widehat{y} - y)^2\right] + 0 + E\left[\sum_{i=1}^{n} w_i^2 \varepsilon_i^2\right]$$

($\because$ ε_i is independent of ε_j and ε_i is independent of ($\widehat{y}$-y))

$$= (E\left[(\widehat{y} - y)^2\right] + \sigma^2 \sum_{i=1}^{n} w_i^2$$

(same as L_2 norm penalty)