College of Engineering & Applied Science
UNIVERSITY OF COLORADO **BOULDER**

University of Colorado Boulder **Amulya Ambati**
CSCI 5922: Neural Nets and Deep Learning **Problem2**
Prof. Danna Gurari **February 14, 2024**

---

1) Model Parameters versus Hyperparameters

Model parameters are variables in the model that are learned by the model upon training on a dataset. Whereas Hyperparameters are those variables that need to be explicitly mentioned to conduct training. Model parameters are required for making predictions and are learned in the process of training. Hyper-parameters drive the training process and are used for optimizing training and in turn decide the model parameters.
A particular model architecture can be trained by providing various sets of hyperparameters. We generally select the best hyperparameters based on their performance on validation data rather than on training data, because it reflects the model's performance on unseen datasets. Hence the final set of hyperparameters is generally selected by a human upon comparing the validation performance.

2) Overfitting vs Underfitting

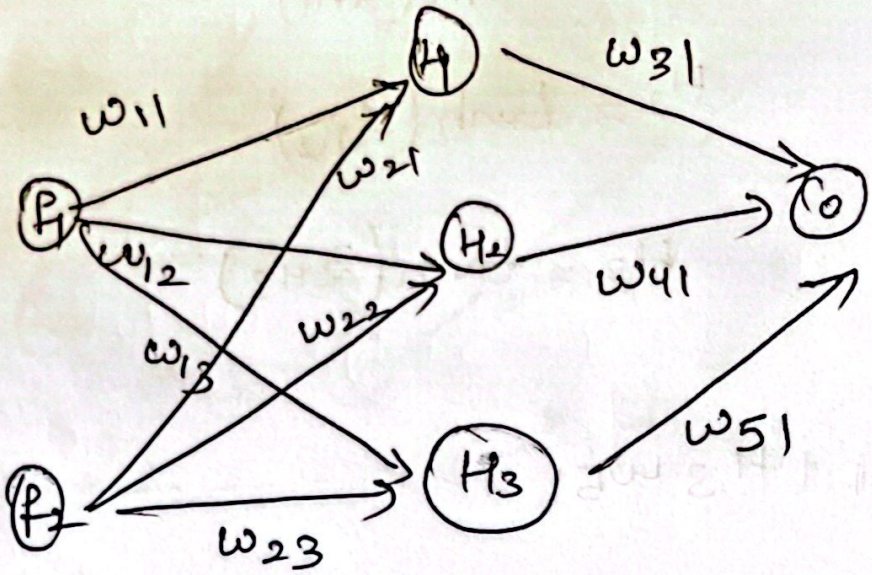
a) If a model performs poorly on both training and new examples that means the model is underfitting. Underfitting is a case when the model fails to even learn and memorize the train data and has poor performance on it and thereby obviously cannot generalize and performs poorly on new examples as well. Some of the reasons that cause underfitting are:

- Simple model architecture on complex dataset.
- Inadequate amount of training data.
- Excessive regularization might constrain the model to capture the data well.

b) If a model performs well on training data and poorly on new examples that means the model is overfitting. Overfitting is a case when the model captures all the nuances in the dataset, training data's noise, and random fluctuations rather than picking on the underlying generalized patterns which thereby causes the model to have high performance on train data and cannot generalize on unseen examples. Some of the reasons that cause overfitting are:

- Too complex model architecture.
- Lots of noise in the training data.
- Poor regularization.

3)



FP $z_{H1} = f_1 w_{11} + f_2 w_{21} + \beta_1 \qquad H_1 = \tanh(z_{H1})$

$z_{H2} = f_1 w_{12} + f_2 w_{22} + \beta_2 \qquad H_2 = \tanh(z_{H2})$

$z_{H3} = f_1 w_{13} + f_3 w_{23} + \beta_3 \qquad H_3 = \tanh(z_{H3})$

$$o = H_1 w_{31} + H_2 w_{41} + H_3 w_{51} + \beta_4$$

$$y = \frac{1}{1+e^{-o}}$$

BP $$L_{CE}(y,t) = -t\log(y) - (1-t)\log(1-y)$$

$$\frac{d}{dy} L_{CE} = \frac{y-t}{y(1-y)}$$

→ ~~$\frac{d}{dy} w_{31} = \frac{dL}{d}$~~

$$\frac{d}{dw_{31}} L_{CE} = \frac{dL_{CE}}{dy} \times \frac{dy}{do} \times \frac{do}{dw_{31}}$$

$$= \frac{y-t}{y(1-y)} \times \frac{e^{-o}}{(1+e^{-o})^2} \times H_1$$

$$\frac{d}{dw_{41}} L_{CE} = \frac{y-t}{y(1-y)} \times \frac{e^{-o}}{(1+e^{-o})^2} \times H_2$$

$$\frac{d}{dw_{51}} L_{CE} = \frac{y-t}{y(1-y)} \times \frac{e^{-o}}{(1+e^{-o})^2} \times H_3$$

$$\frac{d}{dw_{11}} L_{CE} = \frac{dL_{CE}}{dy} \times \frac{dy}{do} \times \frac{do}{dH_1} \times \frac{dH_1}{dz_{H1}} \times \frac{dz_{H1}}{dw_{11}}$$

$$= \frac{y-t}{y(1-y)} \times \frac{e^{-o}}{(1+e^{-o})^2} \times w_{31} \times 4 \times \frac{e^{-2z_{H1}}}{(1+e^{-2z_{H1}})^2} \times f_1$$

$$\frac{d}{dx}\tanh x = \frac{d}{dx}(2\sigma(2x)-1)$$

$$= 2\frac{d}{dx}\sigma(2x)$$

$$= 4 \times \sigma'(2x)$$

$$\frac{d L_{CE}}{d\beta_4} = \frac{dL_{CE}}{dy} \times \frac{dy}{do} \times \frac{do}{d\beta_4}$$

$$= \frac{y-t}{y(1-y)} \times \frac{e^{-o}}{(1+e^{-o})^2} \times 1$$

$$\frac{d}{dw_{21}} L_{CE} = \frac{y-t}{y(1-y)} \times \frac{e^{-o}}{(1+e^{-o})^2} \times w_{31} \times 4 \times \frac{e^{-2\beta_{H1}}}{(1+e^{-2\beta_{H1}})^2} \times f_2$$

$$\frac{d}{dw_{12}} L_{CE} = \frac{y-t}{y(1-y)} \times \frac{e^{-o}}{(1+e^{-o})^2} \times w_{41} \times 4 \times \frac{e^{-2\beta_{H2}}}{(1+e^{-2\beta_{H2}})^2} \times f_1$$

$$\frac{d}{dw_{22}} L_{CE} = \frac{y-t}{y(1-y)} \times \frac{e^{-o}}{(1+e^{-o})^2} \times w_{41} \times 4 \times \frac{e^{-2\beta_{H2}}}{(1+e^{-2\beta_{H2}})^2} \times f_2$$

$$\frac{d}{dw_{13}} L_{CE} = \frac{y-t}{y(1-y)} \times \frac{e^{-o}}{(1+e^{-o})^2} \times w_{51} \times 4 \times \frac{e^{-2\beta_{H3}}}{(1+e^{-2\beta_{H3}})^2} \times f_1$$

$$\frac{d}{dw_{23}} = \frac{y-t}{y(1-y)} \times \frac{e^{-o}}{(1+e^{-o})^2} \times w_{51} \times 4 \times \frac{e^{-2\beta_{H3}}}{(1+e^{-2\beta_{H3}})^2} \times f_2$$

→ Substitute Values

$f_1 = 100$, $f_2 = 20$, $y = 1$

$w_{11} = -0.1$, $w_{21} = 0.05$

$\beta_1 = 0$

$w_{12} = 0.1$, $w_{22} = -0.1$

$\beta_2 = 0$

$w_{13} = -0.2$, $w_{23} = 0.15$

$\beta_3 = 0$

$w_{31} = 1$

$w_{41} = 0.5$

$w_{51} = 0.1$

$\beta_4 = 0$

FP $z_{H1} = 100(-0.1) + 20(0.05) + 0 = -10 + 1$

$= -9$

$z_{H2} = 100(0.1) + 20(-0.1) + 0 = 10 - 2 = 8$

$z_{H3} = 100(-0.2) + 20(0.15) + 0 = -20 + 3 = -17$

$H_1 = \tanh(-9), \quad H_2 = \tanh(8), \quad H_3 = \tanh(-17)$

$0 = H_1\left(\frac{-0.2}{1}\right) + H_2(0.5) + H_3(0.1) + 0$

$= \tanh(-9) + 0.5\tanh(8) + 0.1\tanh(-17)$

$= -1 + 0.5(1) + 0.1(-1)$

$= -0.6$

$$\hat{y} = \frac{1}{1+e^{-0}} = \frac{1}{1+e^{0.6}} = 0.354$$

BP $$\frac{d}{dy} L_{CE} = \frac{(0.354 - 1)}{0.354\,(1 - 0.354)} = \frac{-1}{0.354}$$

$$\frac{d}{dw_{31}} L_{CE} = \frac{-1}{0.354} \times \frac{e^{0.6}}{(1+e^{0.6})^2} \times \tanh(-9)$$

$$= \frac{-1}{0.354} \times 0.229 \times -1$$

$$= 0.646$$

$$\frac{e^{0.6}}{(1+e^{0.6})^2} = 0.229$$

$$w_{31} = w_{31} - \alpha \frac{d}{dw_{31}} L_{CE}$$

$$= 1 - 0.1(0.646) = +0.935$$

$$\frac{d\,L_{CE}}{dw_{41}} = \frac{-1}{0.354} \times 0.229 \times \tanh(8) = \frac{-1}{0.354} \times 0.229 \times 1$$

$$= -0.646$$

$$w_{41} = w_{41} - \alpha\left(\frac{d\,L_{CE}}{dw_{41}}\right) = 0.5 - 0.1(-0.646)$$

$$= 0.565$$

$$\frac{d\,L_{CE}}{dw_{51}} = \frac{-1}{0.354} \times 0.229 \times \tanh(-17) = \frac{-1}{0.354} \times 0.229 \times -1$$

$$= 0.646$$

$$w_{51} = w_{51} - \alpha\left(\frac{d\,L_{CE}}{dw_{51}}\right) = 0.1 - 0.1(0.646)$$

$$= +0.035$$

$$\frac{d\,L_{CE}}{d\beta_4} = \frac{-1}{0.354} \times 0.229 = -0.646$$

$$\beta_4 = \beta_4 - \alpha\left(\frac{d\,L_{CE}}{d\beta_4}\right) = 0 - (-0.646)$$

$$= 0.646$$

$$\frac{dLCE}{dw_{11}} = \frac{-1}{0.354} \times 0.229 \times 1 \times 4 \times \frac{e^{-2(-9)}}{(1+e^{-2(-9)})^2} \times 100$$

$$= -0.00003$$

$$\cong 0$$

$$w_{11} = w_{11} - \alpha\left(\frac{dLCE}{dw_{11}}\right) = w_{11} = -0.1$$

$$\frac{dLCE}{dw_{21}} = \frac{-1}{0.354} \times 0.229 \times 1 \times 4 \times \frac{e^{-2(-9)}}{1+(e^{-2(-9)})^2} \times 20$$

$$\cong 0$$

$$w_{21} = w_{21} - \alpha\left(\frac{dLCE}{dw_{21}}\right) = w_{21} = 0.05$$

$$\frac{dLCE}{d\beta_1} = \frac{-1}{0.354} \times 0.229 \times 1 \times 4 \times \frac{e^{-2(-9)}}{1+(e^{-2(-9)})^2} \times 1$$

$$\cong 0$$

$$\beta_1 = \beta_1 - \alpha\left(\frac{dLCE}{d\beta_1}\right) = \beta_1 = 0$$

$$\frac{dLCE}{dw_{12}} = \frac{-1}{0.354} \times 0.229 \times 0.5 \times 4 \times \frac{e^{-2(8)}}{(1+e^{-2(8)})^2} \times 100$$

$$= 0.000014$$

$$\cong 0$$

$$w_{12} = w_{12} - \alpha\left(\frac{dLCE}{dw_{12}}\right) = w_{12} = 0.1$$

$$\frac{dLCE}{dw_{22}} = \frac{-1}{0.354} \times 0.229 \times 0.5 \times 4 \times \frac{e^{-2(8)}}{(1+e^{-2(8)})^2} \times 20 \cong 0$$

$$w_{22} = w_{22} - \alpha\left(\frac{dLCE}{dw_{22}}\right) = w_{22} = -0.1$$

$$\frac{dLCE}{d\beta_2} = \frac{-1}{0.354} \times 0.229 \times 0.5 \times 4 \times \frac{e^{-2(8)}}{(1+e^{-2(8)})^2}$$

$$\simeq 0$$

$$\beta_2 = \beta_2 - \alpha\left(\frac{dLCE}{d\beta_2}\right) \simeq \beta_2 = 0$$

$$\frac{dLCE}{dw_{13}} = \frac{-1}{0.354} \times 0.229 \times 0.1 \times 4 \times \frac{e^{-2(-17)}}{(1+e^{-2(-17)})^2} \times 100$$

$$\simeq 0$$

$$w_{13} = w_{13} - \alpha\left(\frac{dLCE}{dw_{13}}\right) = w_{13} = -0.12$$

$$\frac{dLCE}{dw_{23}} = \frac{-1}{0.354} \times 0.229 \times 0.1 \times 4 \times \frac{e^{-2(-17)}}{(1+e^{-2(-17)})^2} \times 20$$

$$\simeq 0$$

$$w_{23} = w_{23} - \alpha\left(\frac{dLCE}{dw_{23}}\right) = w_{23} = 0.15$$

$$\frac{dLCE}{d\beta_3} = \frac{-1}{0.354} \times 0.229 \times 0.1 \times 4 \times \frac{e^{-2(-17)}}{(1+e^{-2(-17)})^2} \times 1$$

$$\simeq 0$$

$$\beta_3 = \beta_3 - \alpha\left(\frac{dLCE}{d\beta_3}\right) = \beta_3 = 0 \Longrightarrow$$

Post Sample 1

w_{11}	w_{21}	β_1	w_{12}	w_{22}	β_2	w_{13}	w_{23}	β_3
-0.1	0.05	0	0.1	-0.1	0	-0.2	0.15	0

w_{31}	w_{41}	w_{51}	β_4
0.935	0.565	0.035	0.646

Update with Sample 2 :

$f_1 = 25, \quad f_2 = 4, \quad y = 0$

FP

$z_{H1} = 25(-0.1) + 4(0.05) + 0 = -2.3$

$z_{H2} = 25(0.1) + 4(-0.1) + 0 = 2.1$

$z_{H3} = 25(-0.2) + 4(0.15) + 0 = -4.4$

$H_1 = \tanh(-2.3), \quad H_2 = \tanh(2.1), \quad H_3 = \tanh(-4.4)$

$0 = H_1 w_{31} + H_2 w_{41} + H_3 w_{51} + \beta_4$

$= 0.935(-0.98) + 0.565(0.970) + 0.035(-0.999) + 0.646$

$= 0.243$

$$\hat{y} = \frac{1}{1+e^{-0}} = \frac{1}{1+e^{-0.243}} = 0.560$$

BP

$$\frac{d_{LCE}}{dy} = \frac{(0.560-0)}{0.560(1-0.560)} = \frac{1}{0.44} = 2.273$$

$$\frac{dLCE}{dw_{31}} = 2.273 \times \frac{e^{-0.243}}{(1+e^{-0.243})^2} \times \tanh(-2.3)$$

$$= 2.273 \times 0.246 \times (-0.98)$$

$$= -0.548$$

$$\frac{e^{-0.243}}{(1+e^{-0.243})} = 0.246$$

$$w_{31} = w_{31} - \alpha\left(\frac{dLCE}{dw_{31}}\right)$$

$$= 0.935 - 0.1(-0.548) = 0.989$$

$$\frac{dLCE}{dw_{41}} = 2.273 \times 0.246 \times \tanh(2.1)$$

$$= 0.542$$

$$w_{41} = w_{41} - \alpha\left(\frac{dLCE}{dw_{41}}\right) = 0.565 - 0.1(0.542)$$

$$= 0.511$$

$$\frac{dLCE}{dw_{51}} = 2.273 \times 0.246 \times \tanh(-4.4)$$

$$= -0.559$$

$$w_{51} = w_{51} - \alpha\left(\frac{dLCE}{dw_{51}}\right) = 0.035 - (0.1)(-0.559)$$

$$= +0.091$$

$$\frac{dLCE}{dB_4} = 2.273 \times 0.246 = 0.560$$

$$\beta_4 = \beta_4 - \alpha\left(\frac{dLCE}{d\beta_4}\right) = 0.646 - 0.1(0.560)$$

$$= +0.590$$

$$\frac{dLCE}{dw_{11}} = 2.273 \times 0.246 \times 0.935 \times \frac{e^{-2(-2.3)}}{\left(1+e^{-2(-2.3)}\right)^2} \times 25 \times 4$$

$$= 0.129 \times 4 = 0.516$$

$$w_{11} = w_{11} - \alpha\left(\frac{dLCE}{dw_{11}}\right) = -0.1 - 0.1(0.129 \times 4)$$

$$= -0.113$$

$$= -0.151$$

$$\frac{e^{-2(-2.3)}}{\left(1+e^{-2(-2.3)}\right)} = 0.0098$$

$$\frac{dLCE}{dw_{21}} = 2.273 \times 0.246 \times 0.935 \times 4 \times 0.0098 \times 4$$

$$= 0.082$$

$$w_{21} = w_{21} - \alpha\left(\frac{dLCE}{dw_{21}}\right) = 0.05 - 0.1(0.082)$$

$$= +0.042$$

$$\frac{dLCE}{d\beta_1} = 2.273 \times 0.246 \times 0.935 \times 4 \times 0.0098 \times 1$$

$$= 0.020$$

$$\beta_1 = \beta_1 - \alpha\left(\frac{dLCE}{d\beta_1}\right) = 0 - 0.1(0.020) = -0.002$$

$$\frac{dLCE}{dw_{12}} = 2.273 \times 0.246 \times 0.565 \times 4 \times \frac{e^{-2(2.1)}}{\left(1+e^{-2(2.1)}\right)^2} \times 25$$

$$= 2.273 \times 0.246 \times 0.565 \times 4 \times 0.014 \times 25$$

$$= 0.460$$

$$w_{12} = w_{12} - \alpha\left(\frac{dLCE}{dw_{12}}\right) = 0.1 - 0.1(0.460)$$

$$= 0.054$$

$$\frac{e^{-2(2.1)}}{\left(1+e^{-2(2.1)}\right)^2} = 0.014$$

$$\frac{dLCE}{dw_{22}} = 2.273 \times 0.246 \times 0.565 \times 4 \times 0.014 \times 4$$

$$= 0.071$$

$$w_{22} = w_{22} - \alpha\left(\frac{dLCE}{dw_{22}}\right) = -0.1 - 0.1(0.071)$$

$$= -0.107$$

$$\frac{dLCE}{d\beta_2} = 2.273 \times 0.246 \times 0.565 \times 0.014 \times 4$$

$$= 0.006$$

$$\beta_2 = \beta_2 - \alpha\left(\frac{dLCE}{d\beta_2}\right) = 0 - 0.1(0.006)$$

$$= -0.0006$$

$$\frac{dLCE}{dw_{13}} = 2.273 \times 0.246 \times 0.035 \times 4 \times \frac{e^{-2(-4.4)}}{(1+e^{-2(-4.4)})^2} \times 25$$

$$= 2.273 \times 0.246 \times 0.035 \times 4 \times 0.0015 \times 25$$

$$= 0.00001 \times 25$$

$$= 0.00029$$

$$\frac{e^{-2(-4.4)}}{(1+e^{-2(-4.4)})} = 0.0015$$

$$w_{13} = w_{13} - \alpha\left(\frac{dLCE}{dw_{13}}\right) = -0.2 - 0.1(0.0029)$$

$$= -0.20003$$

$$\frac{dLCE}{dw_{23}} = 2.273 \times 0.246 \times 0.035 \times 4 \times 0.015 \times 4$$

$$= 0.00001 \times 4$$

$$= 0.00004$$

$$w_{23} = w_{23} - \alpha\left(\frac{dLCE}{dw_{23}}\right) = 0.15 - 0.1(0.0004) = 0.149$$

$$\frac{dLCE}{d\beta_3} = 0.00001$$

$$\beta_3 = \beta_3 - \alpha\left(\frac{dLCE}{d\beta_3}\right) = 0 - 0.1(0.00001)$$
$$\simeq 0$$

Post sample 2

w_{11} -0.151	w_{21} 0.042	β_1 -0.002	w_{12} 0.054	w_{22} -0.107	β_2 -0.0016
w_{13} -0.20003	w_{23} 0.149	β_3 0	w_{31} $+0.989$	w_{41} 0.511	w_{51} $\neq 0.091$
β_4 $+0.590$					