Exercise 6

Malik Marden

1.

a) Underfitting happens when a model is too simple or not trained well enough, so it fails to capture the underlying structure in the training data. As a result, both training error and testing error stay high, because the model never really learned the task in the first place.

Overfitting happens when a model is too complex or trained for too long, so it starts memorizing the training data instead of learning general patterns. This leads to very low training error, but the test error increases because the model performs poorly on unseen data.

b)
The bias-variance tradeoff describes how model complexity affects prediction error. A high bias means that the model is too simple and underfits and a high variance means means the the model is too complex and overfits. So reducing one increases the other. The goal is to find the right balance that gives the best generalization.

c)
Model 1: Overfitting, because very low training error and test error goes back up.
Model 2: Underfitting, because test and train error stay high, so model doesn't learn pattern.
Model 3: Good fit, Both erros decrease and stay close, which shows good generalization.

2.

Critical Point:

$$\nabla f(x_1, x_2) = \begin{pmatrix} 2x_1 + 2x_2 \\ 2x_1 - 2x_2 \end{pmatrix} = 0$$

$2x_1 + 2x_2 = 0$

$x_1 = -x_2$

$\hookrightarrow$ substitute

$x_1 = -x_1$

$2x_1 = 0$

$x_1 = 0$

$2x_1 - 2x_2 = 0$

$x_1 = x_2$

$\hookrightarrow$ substitute

$x_2 = -x_2$

$2x_2 = 0$

$x_2 = 0$

Critical point $(0,0)$.

Hessian Matrix:

$$H = \begin{bmatrix} \frac{d^2 f}{dx_1^2} & \frac{d^2 f}{dx_1 x_2} \\ \frac{d^2 f}{dx_2 x_1} & \frac{d^2 f}{dx_2^2} \end{bmatrix}$$

$\frac{df}{dx_1} = 2x_1 + 2x_2$ $\quad \frac{df}{dx_2} = 2x_1 - 2x_2$

$$H = \begin{pmatrix} 2 & 2 \\ 2 & -2 \end{pmatrix}$$

Eigenvectors and Eigenvalues of H:

$\det(H-\lambda I)=0$

$\det\begin{pmatrix} a & b \\ c & d \end{pmatrix} = ad-bc$

$H-\lambda I = \begin{pmatrix} 2-\lambda & 2 \\ 2 & -2-\lambda \end{pmatrix}$

$(2-\lambda)\cdot(-2-\lambda) - 2\cdot 2 = -4+\lambda^2-4 = \lambda^2-8$

$\lambda^2-8=0$

$=\lambda^2=8 \quad \Rightarrow \quad \lambda_1=\sqrt{8} \quad x_2=-\sqrt{8}$

$\det(H-\lambda I)v=0$

λ_1:

$H-\sqrt{8}I = \begin{pmatrix} 2-\sqrt{8} & 2 \\ 2 & -2-\sqrt{8} \end{pmatrix}$

$(2-\sqrt{8})v_1 + 2v_2 = 0$

$(2-\sqrt{8})v_1 = -2v_2$

$v_1 = -\frac{2v_2}{2-\sqrt{8}} = (1+\sqrt{2})v_2 \qquad v_2=1$

$v_1\begin{pmatrix} 1+\sqrt{2} \\ 1 \end{pmatrix}$

λ_2:

$$|H + \sqrt{8}\,I| = \begin{pmatrix} 2+\sqrt{8} & 2 \\ 2 & -2+\sqrt{8} \end{pmatrix}$$

$$(2+\sqrt{8})v_1 + 2v_2 = 0 \quad | -2v_2$$

$$(2+\sqrt{8})v_1 = -2v_2 \quad | : (2+\sqrt{8})$$

$$v_1 = \frac{-2}{2+\sqrt{8}} v_2 = (1-\sqrt{2})\, v_2 \qquad v_2 = 1$$

$$v_1 \begin{pmatrix} 1-\sqrt{2} \\ 1 \end{pmatrix}$$

Since the Hessian has one positive and one negative eigenvalue, the critical point $(0,0)$ is a saddle point. The surface is convex along the eigenvector with positive eigenvalues and concave along the eigenvector with negative eigenvalue