2.

$$
\begin{aligned}
L[\phi] &= \sum_{i=1}^{I} \ell_i = \sum_{i=1}^{I} \left(\mathrm{f}[x_i, \phi] - y_i\right)^2 \\
&= \sum_{i=1}^{I} \left(\phi_0 + \phi_1 x_i - y_i\right)^2, \qquad (6.5)
\end{aligned}
$$

$$\frac{\partial L}{\partial \phi_0} = \sum_{i=1}^{I} 2(\phi_1 x_i - y_i)$$

$$\frac{\partial^2 L}{\partial \phi_0^2} = 0$$

$$\frac{\partial L}{\partial \phi_1} = \sum_{i=1}^{I} 2x_i(\phi_0 + \phi_1 x_i - y_i)$$

$$\frac{\partial^2 L}{\partial \phi_1^2} = \sum_{i=1}^{I} 2x_i^2$$

$$\frac{\partial^2 L}{\partial \phi_0 \partial \phi_1} = \sum_{i=1}^{I} 2x_i$$

$$\mathrm{Tr}(\underline{\underline{H}}) = 0 + \sum_{i=1}^{I} 2x_i^2 > 0$$

$$\mathrm{Det}(\underline{\underline{H}}) = \left(\sum_{i=1}^{I} 2x_i\right)^2 - 0 > 0$$

$\therefore$ The function is convex.

4. Binary cross-entropy.

$$Pr(y=1|x) = \mathrm{sig}[\phi_0 + \phi_1 x]$$
$$Pr(y=0|x) = 1 - \mathrm{sig}[\phi_0 + \phi_1 x]$$

i) this is a sigmoid function where the x-shift is determined by ϕ_0 and the steepness is determined by ϕ_1

ii) $Pr(y|x) = \mathrm{sig}[\phi_0 + \phi_1 x]^y \cdot (1 - \mathrm{sig}[\phi_0 + \phi_1 x]^{1-y})$

$$L[\phi] = -\sum_{i}^{I}\left[y_i \log[\mathrm{sig}[\phi_0 + \phi_1 x]] + (1-y_i)\log[1 - \mathrm{sig}[\phi_0 + \phi_1 x]]\right]$$

iii) $\frac{\partial \, sig(z)}{\partial z} = \frac{e^{-z}}{(1+e^{-z})^2}$ $\qquad sig = \frac{1}{1+e^{-z}}$

$$\frac{\partial L}{\partial \phi_0} = \sum_i^I \frac{(1-y_i)}{(1-sig[\phi_0+\phi_1 x_i])} \frac{e^{-(\phi_0+\phi_1 x_i)}}{(1+e^{-(\phi_0+\phi_1 x_i)})^2} \frac{-y_i}{(1-sig[\phi_0+\phi_1 x_i])} \frac{e^{-(\phi_0+\phi_1 x_i)}}{(1+e^{-(\phi_0+\phi_1 x_i)})^2}$$

Similar for $\frac{\partial L}{\partial \phi_1}$

6.

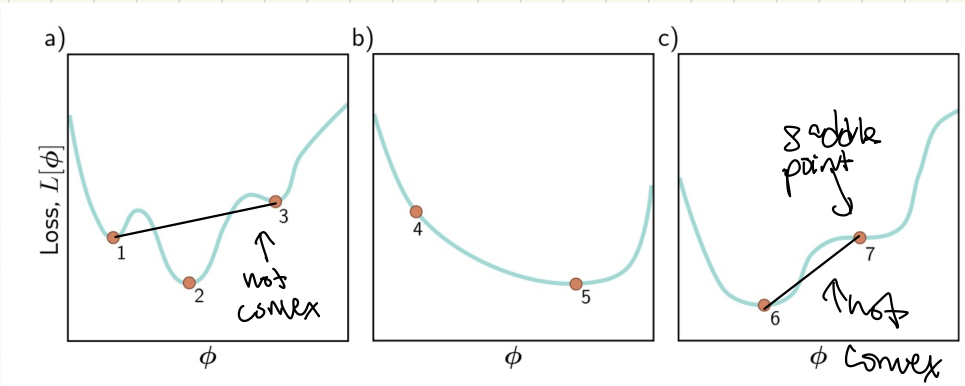


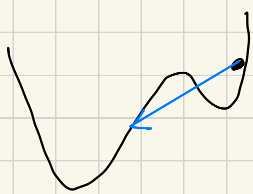
Figure 6.11 Three 1D loss functions for problem 6.6.

Convex: if you draw a line between any two points then the line is either above or on the line.
So only b is convex.

i) 1, 3
ii) 2, 5, 6
iii) 4, 7.

8. Yes, if the gradient is big enough

e.g.



6. $m_0 = 0$

$$m_1 = \beta m_0 + (1-\beta)\frac{\partial L(\Phi_0)}{\partial \Phi} = (1-\beta)\frac{\partial L(\Phi_0)}{\partial \Phi}$$

$$m_2 = \beta(m_1) + (1-\beta)\frac{\partial L(\Phi_1)}{\partial \Phi_0}$$

$$= \beta(1-\beta)\frac{\partial L(\Phi_0)}{\partial \Phi} + (1-\beta)\frac{\partial L(\Phi_1)}{\partial \Phi}$$

$$m_3 = \beta^2(1-\beta)\frac{\partial L(\Phi_0)}{\partial \Phi} + \beta(1-\beta)\frac{\partial L(\Phi_1)}{\partial \Phi}$$

$$m_T = \sum_{t=0}^{T-1} \beta^{T-t-1}(1-\beta)\frac{\partial L(\Phi_t)}{\partial \Phi}$$