$x = [x_1\ x_2 \cdots x_N]$: feature $\in \mathbb{R}^N$

$W = [w_1\ w_2 \cdots w_N]$: weight $\in \mathbb{R}^N$

$a = W \cdot x$: logit $\in \mathbb{R}$

$P_\theta(\text{positive} \mid x) = \sigma(a) \in \mathbb{R}$

g: ground truth $\in \{0, 1\}$

- loss

$$L = -g \log P_\theta - (1-g)\log(1-P_\theta)$$

- goal

show hessian$(L) \geq 0$

(assume $g=1$ below)

$$\frac{\partial L}{\partial w_i} = \underbrace{\frac{\partial P}{\partial w_i}}_{(1)} \underbrace{\frac{\partial L}{\partial P}}_{(2)}$$

$$(1): \frac{\partial P}{\partial w_i} = \frac{\partial a}{\partial w_i}\frac{\partial P}{\partial a} = x_i P(1-P)$$

$$(2): \frac{\partial L}{\partial P} = -\frac{1}{P}$$

$$\therefore \frac{\partial L}{\partial w_i} = x_i P(1-P)\left(-\frac{1}{P}\right)$$

$$= -x_i(1-P)$$

$$\frac{\partial^2 L}{\partial w_j \partial w_i} = \frac{\partial}{\partial w_j}\left(-x_i(1-P)\right)$$

$$= x_i x_j P(1-P) \leftarrow \text{reuse } (1)$$

$$H(L) = P(1-P)\begin{pmatrix} x_1^2 & x_1x_2 & \cdots & x_1x_N \\ x_2x_1 & x_2^2 & \cdots & x_2x_N \\ & \vdots & & \\ x_Nx_1 & \cdots & \cdots & x_N^2 \end{pmatrix}$$

$$= P(1-P)\begin{pmatrix} x_1 x \\ x_2 x \\ \vdots \\ x_N x \end{pmatrix} \in \mathbb{R}^{N \times N}$$

v : a non-zero vector $\in \mathbb{R}^N$

$$H v^T = P(1-P)(x \cdot v)\, x \in \mathbb{R}^N$$

$$v H v^T = P(1-P)(x \cdot v)^2 \geq 0$$

$\rightarrow$ convex