

Practical Class 04

Convex Functions and Logistic Regression

1. Consider the sigmoid function $\sigma(x) = \frac{1}{1+\exp(-x)}$. Is this a convex function? Show your working.

Solution: Recall that if f is twice differentiable, convexity is the same as

$$f''(x) \geq 0; \forall x \in \text{Dom}(f).$$

The answer is NO and we check the second order derivative as follows: Apply the quotient rule of derivatives,

$$\frac{d}{dx}\sigma(x) = \frac{-(-1)\exp(-x)}{(1+\exp(-x))^2} = \frac{\exp(-x)}{1+\exp(-x)} \times \frac{1}{1+\exp(-x)}$$

Notice that we can rewrite the first term as

$$\frac{\exp(-x)}{1+\exp(-x)} = \frac{(1+\exp(-x)) - 1}{1+\exp(-x)} = 1 - \frac{1}{1+\exp(-x)},$$

together with the definition of sigmoid function, we have

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

Then use the product rule of derivatives, we have

$$\begin{aligned}\frac{d^2}{dx^2}\sigma(x) &= \left(-\frac{d}{dx}\sigma(x)\right)\sigma(x) + (1 - \sigma(x))\frac{d}{dx}\sigma(x) \\ &= -(1 - \sigma(x))\sigma(x)\sigma(x) + (1 - \sigma(x))(1 - \sigma(x))\sigma(x) \\ &= (1 - \sigma(x))\sigma(x)(1 - 2\sigma(x))\end{aligned}$$

Notice that $\exp(x) > 0$ for all $x \in \mathbb{R}$, therefore both $\sigma(x)$ and $1 - \sigma(x)$ are positive for all x . But $1 - 2\sigma(x)$ could be negative for some values of x and therefore the second derivatives is not always positive, therefore $\sigma(x)$ is NOT a convex function.

2. Consider again the sigmoid function $\sigma(x) = \frac{1}{1+\exp(-x)}$. Is $-\log(\sigma(x))$ a convex function? Show your working.

Solution: Using the properties of log functions, we have

$$-\log(\sigma(x)) = \log\left(\frac{1}{1+\exp(-x)}\right)^{-1} = \log(1+\exp(-x)),$$

thus

$$\begin{aligned}\frac{d}{dx}(-\log(\sigma(x))) &= \frac{-\exp(x)}{1 + \exp(-x)} \\ &= \frac{-1 - \exp(x) + 1}{1 + \exp(-x)} = \frac{-(1 + \exp(-x)) + 1}{1 + \exp(-x)} \\ &= -1 + \frac{1}{1 + \exp(-x)} = \sigma(x) - 1.\end{aligned}$$

Using results from previous questions, we have

$$\frac{d^2}{dx^2}(-\log(\sigma(x))) = \frac{d}{dx}\sigma(x) = (1 - \sigma(x))\sigma(x) > 0.$$

Therefore $-\log(\sigma(x))$ is a convex function.

3. A function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ in the form

$$f(\mathbf{x}) = f(x_1, \dots, x_n) = a_0 + a_1x_1 + \dots + a_nx_n = \mathbf{a}^T \mathbf{x} + a_0$$

is called an affine function. An affine function is convex, convince yourself that this is the case. (In fact, affine functions are the only functions that are both convex and concave as $f''(\mathbf{x}) = 0$.)

Now, let $g : \mathbb{R} \rightarrow \mathbb{R}$ be a convex function. Is that $g \circ f = g(f(\mathbf{x}))$ also convex? Show your working.

Solution: By the chain rule of derivatives,

$$\frac{d}{d\mathbf{x}}g(f(\mathbf{x})) = g'(f(\mathbf{x}))f'(\mathbf{x}).$$

Apply the product rule and chain rule again,

$$\frac{d^2}{d\mathbf{x}^2}g(f(\mathbf{x})) = [g''(f(\mathbf{x}))f'(\mathbf{x})]f'(\mathbf{x}) + g'(f(\mathbf{x}))f''(\mathbf{x}) = g''(f(\mathbf{x}))[f'(\mathbf{x})]^2 + g'(f(\mathbf{x}))f''(\mathbf{x}).$$

The second term is 0 as $f''(\mathbf{x}) = 0$. The first term is non negative as g is convex and the square term is non-negative. Therefore, the composition $g \circ f$ is also a convex function.

4. Use all the results above, show that the loss of a logistic model, which is written in the form of negative log-likelihood, is convex.

Solution: The loss (negative log-likelihood) is

$$L(\theta) = -\sum_{i=1}^n [y_i \log(\sigma(\theta^T \mathbf{x}_i)) + (1 - y_i)(\log(1 - \sigma(\theta^T \mathbf{x}_i))].$$

First we note that, due to the linearity of differentiation, if f and g are both convex, then $f + g$ is also convex. Also notice that $y_i \in \{0, 1\}$. As such, we only need to show the following two statements:

1. if $y_i = 1$, $-\log(\sigma(\theta^T \mathbf{x}_i))$ is convex in θ ; and
2. if $y_i = 0$, $-\log(1 - \sigma(\theta^T \mathbf{x}_i))$ is convex in θ .

For statement 1, $\theta^T \mathbf{x}_i$ is obviously an affine function and $-\log(\sigma(\mathbf{x}))$ is shown to be convex in question 2. Using the result of question 3 that the composition of an affine function and a convex function is also convex, $-\log(\sigma(\theta^T \mathbf{x}_i))$ is convex in θ . That is, statement 1 is true.

For statement 2, similarly, if we can show that $-\log(1 - \sigma(\mathbf{x}))$ is convex, then by the composition of convex function and affine function, statement 2 is true. The convexity of $-\log(1 - \sigma(\mathbf{x}))$ can be shown using the calculation we obtained in question 1 as follows:

$$\frac{d}{dx}[-\log(1 - \sigma(\mathbf{x}))] = -\frac{-\sigma'(\mathbf{x})}{1 - \sigma(\mathbf{x})} = -\frac{-\sigma(\mathbf{x})(1 - \sigma(\mathbf{x}))}{1 - \sigma(\mathbf{x})} = \sigma(\mathbf{x}),$$

and

$$\frac{d^2}{dx^2}[-\log(1 - \sigma(\mathbf{x}))] = \frac{d}{dx}\sigma(\mathbf{x}) = \sigma(\mathbf{x})(1 - \sigma(\mathbf{x})) \geq 0.$$

Thus, we have shown that this loss function is convex and as a consequence, the minimum exists in a certain domain.