# Exercise 04

## 1.

For the Bernoulli model, we have:

$$p(y_i|\boldsymbol{x}_i, \omega) = \sigma(\omega^{\mathrm{T}}\boldsymbol{x}_i)^{y_i}\left(1 - \sigma(\omega^{\mathrm{T}}\boldsymbol{x}_i)\right)^{1-y_i}.$$

The log-likelihood is then given by:

$$\begin{aligned}
L(\omega) &= \log \prod_{i=1}^{N} \sigma(\omega^{\mathrm{T}}\boldsymbol{x}_i)^{y_i}\left(1 - \sigma(\omega^{\mathrm{T}}\boldsymbol{x}_i)\right)^{1-y_i} \\
&= \sum_{i=1}^{N} (y_i \log \sigma(\omega^{\mathrm{T}}\boldsymbol{x}_i) + (1-y_i)\log(1 - \sigma(\omega^{\mathrm{T}}\boldsymbol{x}_i))) \\
&= \sum_{i=1}^{N} (y_i \log \sigma(\omega^{\mathrm{T}}\boldsymbol{x}_i) + (1-y_i)\log \sigma(-\omega^{\mathrm{T}}\boldsymbol{x}_i)) \\
&= \left(\sum_{y_i=1} + \sum_{y_i=0}\right)(y_i \log \sigma(\omega^{\mathrm{T}}\boldsymbol{x}_i) + (1-y_i)\log \sigma(-\omega^{\mathrm{T}}\boldsymbol{x}_i)) \\
&= \sum_{y_i=1} \log \sigma(\omega^{\mathrm{T}}\boldsymbol{x}_i) + \sum_{y_i=0} \log \sigma(-\omega^{\mathrm{T}}\boldsymbol{x}_i).
\end{aligned}$$

## 2.

### (a)

Easy to show by noticing that $-\partial^2 L(\omega) \geq 0$ always holds. (The computation is not easy though)

Convexity is the solid guarantee that we can find the global minimality.

### (b)

This is also trivial, since $\mathcal{L}_{\mathrm{BCE}}(\boldsymbol{\omega})$ is only different from the task by a negative constant $-\frac{1}{N}$, which indicates the equivalence between maximizing binary cross-entropy and minimizing log-likelihood function.

## 3.

### (a)

As $k \to \infty$, $\log \sigma(\omega^{\mathrm{T}}\boldsymbol{x}_i)$ becomes $\log(1)$ which is zero, whereas $\log \sigma(-\omega^{\mathrm{T}}\boldsymbol{x}_i)$ turns into infinity, hence preventing the convergence for linearly separable data.

### (b)

We can introduce regularization term to make the new objective function strictly convex.