Worksheet 6: Convexity and Gradients

Name:
Due November 3, 2023

1. (Properties of convexity) Show that $\mathrm{f} : \mathbb{R} \to \mathbb{R}$ defined by $\mathrm{f(x)} := \log(1 + \exp(\mathrm{x})) = (\sigma(-\mathrm{t}))^{-1}$ is convex.
   Solution:

$$\begin{aligned} f'(x) &= \frac{\exp(x)}{1+\exp(x)} \\ f''(x) &= \frac{\exp(x)(1+\exp(x)) - \exp(x)(\exp(x))}{(1+\exp(x))^2} \\ &= \frac{\exp(x)}{(1+\exp(x))^2} \geq 0 \end{aligned}$$

   Therefore, $f(x)$ is convex

2. (Use the definition) Suppose that $f : \mathbb{R}^{\mathrm{d}} \to \mathbb{R}$ decomposes as composition

$$\begin{tikzcd} \mathbb{R}^{\mathrm{d}} \arrow[r, "\eta"] \arrow[rr, bend left, "\mathrm{f}"] & \mathbb{R} \arrow[r, "\mathrm{g}"] & \mathbb{R} \end{tikzcd}$$

where $\eta(w) = \langle w, x \rangle + y$. Show that if g is convex, then so is f.
Solution: Let $\lambda \in [0, 1]$ and $x_1, x_2 \in \mathbb{R}^d$

$$\begin{aligned} f(\lambda x_1 + (1-\lambda)x_2) &= g \circ \eta(\lambda x_1 + (1-\lambda)x_2) \\ &= g(\langle w, \lambda x_1 + (1-\lambda)x_2 \rangle + y) \\ &= g(\lambda\langle w, x_1\rangle + (1-\lambda)\langle w, x_2\rangle + \lambda y + (1-\lambda)y) \\ &= g\left(\lambda\left[\langle w, x_1\rangle + y\right] + (1-\lambda)\left[\langle w, x_2\rangle + y\right]\right) \\ &\leq \lambda g \circ \eta(x_1) + (1-\lambda) g \circ \eta(x_2) \\ &= \lambda f(x_1) + (1-\lambda) f(x_2) \end{aligned}$$

3. (Use the definition) Show that if $f_1, \ldots, f_k$ are convex, then $g = \sum_{j=1}^{k} \alpha_j f_j$ is convex for $\alpha_j \geq 0$.
   Solution:

$$\begin{aligned} g(\lambda x_1 + (1-\lambda)x_2) &= \sum_{j=1}^{k} \alpha_j f_j(\lambda x_1 + (1-\lambda)x_2) \\ &\leq \sum_{j=1}^{k} \alpha_j[\lambda f(x_1) + (1-\lambda) f(x_2)] \\ &= \lambda \sum_{j=1}^{k} \alpha_j f(x_1) + (1-\lambda) \sum_{j=1}^{k} \alpha_j f(x_2) \\ &= \lambda g(x_1) + (1-\lambda) g(x_2) \end{aligned}$$

4. (Convexity of $\log \mathrm{Reg}$) Conclude that ERM problem for $\mathcal{H} = \{\sigma(\alpha x - \beta) : \alpha, \beta \in \mathbb{R}\}$ and $c_{\tilde{y}}(x,y) = -y\log(\tilde{y}(x)) - (1-y)\log(1-\tilde{y}(x))$ is convex.
Solution:
   (a) The ERM problem is convex iff this cost function is convex by problem 3.
   (b) This cost function is convex iff we still have convexity after restricting to just $\sigma(x)$ by problem 2
   (c) $-y\log(\sigma(x)) = y\log(\frac{1}{\sigma(x)}) = y\log(1+\exp(-x))$ is convex from problem 1 (the negative doesn't matter because it is convex over all of $\mathbb{R}$ and I am assuming $1 \leq y \geq 0$ which I guess it needs to be).
   (d) Similarly, then $-(1-y)\log(1-\sigma(x)) = (1-y)\log(\frac{1+\exp(-x)}{\exp(-x)}) = (1-y)\log(1+\exp(x))$

5. (Computing Gradients) For $\mathcal{H}_n := \left\{\sigma(p(x)) : p \text{ is degree } n \text{ polynomial, } p(x) = \sum_{j=0}^{n} \alpha_j x^j\right\}$, express $\nabla_\alpha e_S\left(\tilde{y}_\alpha\right)$, where $e_S(\tilde{y}) = \frac{1}{m}\sum_{j=1}^{m} c_{\tilde{y}}\left(x_j, y_j\right)$ denotes the empirical (risk) expectation (implicit: $S = \left(\left(x_1, y_1\right), \ldots, \left(x_m, y_m\right)\right)$) of cost $c_{\tilde{y}}(x,y) = (\tilde{y}(x) - y)^2$.
Solution: First note that $\sigma(x)' = \sigma(x)(1-\sigma(x))$
Secondly, let $\mathbf{x_j} = \begin{bmatrix} 1 & x_j & x_j^2 & ... & x_j^n \end{bmatrix}$

$$\begin{aligned}
\nabla_\alpha e_s(\tilde{y}_\alpha) &= \nabla_\alpha \frac{1}{m}\sum_{j=1}^{m} c_{yt}(x_j, y_j) \\
&= \frac{1}{m}\sum_{j=1}^{m} \nabla_\alpha(\tilde{y}_\alpha(x) - y)^2 \\
&= \frac{1}{m}\sum_{j=1}^{m} 2(\tilde{y}_\alpha(x) - y)\nabla_\alpha(\tilde{y}_\alpha(x)) \\
&= \frac{1}{m}\sum_{j=1}^{m} 2(\sigma(p(x)) - y)\sigma(p(x))(1-\sigma(p(x))\mathbf{x_j}
\end{aligned}$$

6. (Computing Gradients) For same $\mathcal{H}_\mathbf{n}$ as above, express $\nabla_\alpha e_S\left(\tilde{y}_\alpha\right)$, this time $c_{\tilde{y}}(x,y) = -y\log(\tilde{y}(x)) - (1-y)\log(1 - \tilde{y}(x))$. **Solution:**

$$\begin{aligned}
\nabla_\alpha e_s(\tilde{y}_\alpha) &= \nabla_\alpha \frac{1}{m}\sum_{j=1}^{m} c_{yt}(x_j, y_j) \\
&= \frac{1}{m}\sum_{j=1}^{m} \nabla_\alpha\left[-y\log(\tilde{y}_\alpha) - (1-y)\log(1-\tilde{y}_\alpha)\right] \\
&= \frac{1}{m}\sum_{j=1}^{m} \left[\frac{1-y}{1-\tilde{y}_\alpha} - \frac{y}{\tilde{y}_\alpha}\right]\sigma(p(x))(1-\sigma(p(x)))\mathbf{x_j}
\end{aligned}$$