This homework is due at **2pm, September 30, 2019**. There are in total 105 points. Your actual grade of this homework will be $\min\{100, \text{points you get}\}$.

Problem 1

Let $f : \mathbb{R}^p \to \mathbb{R}$. Its *gradient* is a p-dimensional vector given by

$$\nabla f(x) := \left(\frac{\partial f}{\partial x^{(1)}}(x), \dots, \frac{\partial f}{\partial x^{(p)}}(x) \right), \quad \forall x \in \mathbb{R}^p,$$

where $x^{(i)}$ denotes the i-th entry of the vector x. Its *Hessian* is a matrix in $\mathbb{R}^{p \times p}$ given by

$$\left[\nabla^2 f(x) \right]^{(i,j)} := \frac{\partial^2 f}{\partial x^{(i)} \partial x^{(j)}}(x), \quad \forall x \in \mathbb{R}^p,$$

for all $1 \le i, j \le p$, where $\left[\nabla^2 f(x) \right]^{(i,j)}$ denotes the (i, j)-th entry of the matrix $\nabla^2 f(x)$.

Let $a \in \mathbb{R}^p$. A machine learning algorithm called *logistic regression* requires minimizing a sum of functions of the form

$$g(x) := \log\left(1 + \mathrm{e}^{-\langle a, x \rangle}\right), \quad \forall x \in \mathbb{R}^p.$$

1. (15 points) **Show that**

$$\nabla g(x) = \frac{-a}{1 + \mathrm{e}^{\langle a, x \rangle}}, \quad \forall x \in \mathbb{R}^p.$$

2. (15 points) **Show that**

$$\nabla^2 g(x) = \frac{\mathrm{e}^{\langle a, x \rangle} a a^{\mathrm{T}}}{\left(1 + \mathrm{e}^{\langle a, x \rangle}\right)^2}, \quad \forall x \in \mathbb{R}^p,$$

where a^{T} denotes the transpose of a.

3. (15 points) Let $A, B \in \mathbb{R}^{p \times p}$. We write $A \ge B$ if and only if $(A - B)$ is positive semi-definite, and $A \le B$ if and only if $B \ge A$. **Show that**

$$0 \le \nabla^2 g(x) \le \frac{\|a\|_2^2}{4} I, \quad \forall x \in \mathbb{R}^p,$$

where I denotes the identity matrix.

Problem 2

Let ξ be a random variable taking values in $\{-1, 1\}$. Define

$$\varphi(\beta) := \log\left(\mathsf{E}\, \mathrm{e}^{\beta \xi}\right), \quad \forall \beta \in \mathbb{R},$$

where $\mathsf{E}\, \mathrm{e}^{\beta \xi}$ denotes the expectation of $\mathrm{e}^{\beta \xi}$.

1. (15 points) **Show that**

$$\varphi''(\beta) = \mathsf{E}\left[\left(\eta_\beta - \mathsf{E}\, \eta_\beta\right)^2 \right], \quad \varphi'''(\beta) = \mathsf{E}\left[\left(\eta_\beta - \mathsf{E}\, \eta_\beta\right)^3 \right], \quad \forall \beta \in \mathbb{R},$$

for some random variable η_β taking values in $\{-1, 1\}$ whose probability distribution depends on β.

Problem 2 continued on the next page...

2. (15 points) **Show that**

$$\varphi''(\gamma) \le e^{2|\gamma-\beta|}\varphi''(\beta), \quad \forall \beta, \gamma \in \mathbb{R}.$$

HINT: By the results above, we have

$$\varphi'''(\beta) \le 2\varphi''(\beta), \quad \forall \beta \in \mathbb{R}.$$

3. (15 points) **Show that**

$$\varphi'(\gamma) \le \varphi'(\beta) + \left[\frac{e^{2(\gamma-\beta)} - 1}{2\left(\gamma-\beta\right)} \right] \varphi''(\beta)(\gamma-\beta), \quad \forall \beta, \gamma \in \mathbb{R} \textbf{ such that } \gamma > \beta.$$

4. (15 points) **Use the results above to prove that**

$$\log\left[\mathsf{E}\, e^{\lambda(\xi - \mathsf{E}\,\xi)}\right] \le \frac{h(2\lambda)}{4} \operatorname{var} \xi, \quad \forall \lambda > 0,$$

where $h(x) := e^x - x - 1$ **and** $\operatorname{var}\xi$ **denotes the variance of** ξ**.** This is essentially *Bennett's inequality.* See, e.g., [**?**, Theorem 2.9] for the details; however, notice we want a proof based on the results above and do not copy the proof in [**?**].

HINT: Compare $\varphi(\lambda)$ and $\varphi(0)$.