Name:

AMATH 515 Homework Set 1

Due: Wednesday Jan 27th, by midnight.

(1) Let $g : \mathbb{R}^m \to \mathbb{R}$ is a twice differentiable function, $A \in \mathbb{R}^{m\times n}$ any matrix, and h is the composition $g(Ax)$, then
  (a) Show that $\nabla h(x) = A^T \nabla g(Ax)$.
  (b) Show that $\nabla^2 h(x) = A^T \nabla^2 g(Ax) A$
  (c) Use the formulas to compute the gradient and hessian of the logistic regression objective:
$$\sum_{i=1}^{n} \log(1 + \exp(\langle a_i, x\rangle)) - b^T Ax$$
  where a_i denote the rows of A.

(2) Show that each of the following functions is convex.
  (a) Indicator function to a convex set: $\delta_C(x) = \begin{cases} 0 & \text{if} \quad x \in C \\ \infty & \text{if} \quad x \notin C. \end{cases}$
  (b) Support function to any set: $\sigma_C(x) = \sup_{c\in C} c^T x$.
  (c) Any norm (see Chapter 1 for definition of a norm).

(3) Convexity and composition rules. Suppose that f and g are $\mathcal{C}^2$ functions from $\mathbb{R}$ to $\mathbb{R}$, with $h = f \circ g$ their composition, defined by $h(x) = f(g(x))$.
  (a) If f and g are convex, show it is possible for h to be nonconvex (give an example). What additional condition ensures the convexity of the composition?
  (b) If f is convex and g is concave, what additional hypothesis that guarantees h is convex?
  (c) Show that if $f : \mathbb{R}^m \to \mathbb{R}$ is convex and $g : \mathbb{R}^n \to \mathbb{R}^m$ affine, then h is convex.
  (d) Show that the following functions are convex:
    (i) Logistic regression objective: $\sum_{i=1}^{n} \log(1 + \exp(\langle a_i, x\rangle)) - b^T Ax$
    (ii) Poisson regression objective: $\sum_{i=1}^{n} \exp(\langle a_i, x\rangle) - b^T Ax$.

(4) A function f is *strictly convex* if
$$f(\lambda x + (1-\lambda)y) < \lambda f(x) + (1-\lambda) f(y), \quad \lambda \in (0, 1).$$
  (a) Give an example of a strictly convex function that does not have a minimizer. Explain why your function is strictly convex.
  (b) Show that a sum of a strictly convex function and a convex function is strictly convex.

(c) Characterize all solutions to the problem

$$\min_x \frac{1}{2}\|Ax - b\|^2$$

(d) Does the objective function below (logistic regularized with an elastic net) have a minimizer? Is it unique? Explain.

$$\min_x \sum_{i=1}^{n} \log(1 + \exp(\langle a_i, x\rangle)) + \lambda(\alpha\|x\|_1 + (1-\alpha)\|x\|^2), \quad \lambda > 0, \alpha \in (0, 1)$$

(5) Lipschitz constants and β-smoothness. Remember that f is β smooth when its gradient is β-Lipschitz continuous.

(a) Find a global bound for β of the least-squares objective $\frac{1}{2}\|Ax - b\|^2$.

(b) Find a global bound for β of the regularized logistic objective

$$\sum_{i=1}^{n} \log(1 + \exp(\langle a_i, x\rangle)) + \frac{\lambda}{2}\|x\|^2.$$

(c) Do the gradients for Poisson regression admit a global Lipschitz constant?

(6) Behavior of steepest descent for logistic vs. poisson regression.

(a) Given the sample (logistic) data set and starter code, implement gradient descent for ℓ_2-regularized logistic regression. Plot (a) the objective value and (b) the norm of the gradient (as a measure of optimality) on two separate figures. For the figure in (b), make sure the y-axis is on a logarithmic scale.

(b) Implement Newton's method for the same problem. Does the method converge? If necessary, use the line search routine provided to scale your updated directly to ensure descent. Add the plots for Newton's method (a) and (b) to your Figures 1 and 2. What do you notice?

(c) Using the sample (Poisson) data and starter code provided, implement gradient descent and Newton's method for ℓ_2-regularized Poisson regression. You may need to use the line search routine for both algorithms. Make the same plots as you did for the logistic regression examples.

(d) What do you notice qualitatively about steepest descent vs. Newton?