Quiz 2

Convex Optimization, 10-725

Due Friday September 27, 2019

Name:

Andrew ID:

Each question is either in true/false format, or multiple choice. For multiple choice, just choose the single best option. In each case, make sure to fill in the box according to the answer you choose (true or false, or the multiple choice option) completely. All questions are worth 1 point.

1. Gradient descent on a convex function with Lipschitz gradient converges (with suitable step sizes) at the rate:
   ☐ a. $O(1/\epsilon^2)$;
   ☐ b. $O(1/\epsilon)$;
   ☐ c. $O(\log(1/\epsilon))$;
   ☐ d. $O(1/\sqrt{\epsilon})$.

2. Consider the backtracking loop:

$$\begin{aligned} &\text{repeat} \quad t = \beta t \\ &\text{until} \quad f(x + tv) \le f(x) + \alpha t \nabla f(x)^T v, \end{aligned}$$

   where v is the descent direction. A larger value of α leads to fewer backtracking iterations until the backtracking exit condition is satisfied.
   ☐ True
   ☐ False

3. Applying Nesterov's acceleration to gradient descent still results in a descent method.
   ☐ True
   ☐ False

4. If $g(x) = f(Ax + b)$, where f is convex, then subgradients of g are defined by:
   ☐ a. $\partial g(x) = \partial f(Ax + b)$;
   ☐ b. $\partial g(x) = A^T \partial f(Ax + b)$;
   ☐ c. $\partial g(x) = A \partial f(Ax + b)$;
   ☐ d. g need not have subgradients, because it need not be convex.

5. The subdifferential of any function f is always a convex set.
   ☐ True
   ☐ False

6. For any function f, a point x minimizes f if and only if $0 \in \partial f(x)$.
   ☐ True
   ☐ False

7. For any differentiable function f, its subdifferential at a point x is the singleton $\{\nabla f(x)\}$.
   ☐ True
   ☐ False

8. Applying the subgradient method requires knowledge of the full subdifferential of the function in question.
   ☐ True
   ☐ False

9. For the problem of logistic regression with a ridge penalty (i.e., squared ℓ_2 penalty on the parameters), gradient descent and subgradient method, both with same fixed step size, will:
☐ a. converge at different rates, with gradient descent being faster;
☐ b. converge at different rates, with gradient descent being slower;
☐ c. converge at the same rate, because they're doing the exact same thing!
☐ d. it depends on the condition number of $X^T X$, with X being the matrix of predictors.

10. The subgradient method is a descent method.
☐ True
☐ False

11. The subgradient method achieves the optimal rate among nonsmooth first-order methods for minimizing convex functions that are Lipschitz.
☐ True
☐ False

12. Proximal gradient descent is most appealing when the prox is cheap, since convergence (in terms of the number of iterations) will be on par with first-order methods.
☐ True
☐ False

13. The proximal operator of $h(x) = \|x\|_1$ is given by:
☐ a. soft-thresholding;
☐ b. hard-thresholding;
☐ c. projection onto the ℓ_1 ball;
☐ d. not known.

14. For the lasso problem, proximal gradient descent, i.e., ISTA, will often converge:
☐ a. about on par with the subgradient method;
☐ b. faster than the subgradient method;
☐ c. slower than the subgradient method;
☐ d. this is an unfair comparison because these two methods have very different computational costs per iteration.

15. For any convex function f and $t > 0$, the proximal operator of f,

$$\mathrm{prox}_{f,t}(x) = \underset{z}{\mathrm{argmin}} \ \frac{1}{2t}\|x - z\|_2^2 + f(z),$$

is well-defined (meaning the above minimization has a unique solution).
☐ True
☐ False

16. Stochastic methods are generally well-suited to an objective that is a sum of a large number of functions.
☐ True
☐ False

17. Taking larger mini-batches in stochastic gradient descent:
☐ a. reduces the variance, at no additional computational expense;
☐ b. reduces the variance, increases the computational cost of an iteration;
☐ c. reduces the bias, reduces the variance;
☐ d. does not change the variance, but improves communication costs.

18. For stochastic gradient descent, step sizes are commonly chosen by backtracking line search.
☐ True
☐ False

19. Stochastic gradient descent on a convex function that is Lipschitz, converges (with suitable step sizes) at the rate:
☐ a. $O(1/\epsilon)$;

☐ b. $O(1/\sqrt{\epsilon})$;
☐ c. $O(1/\epsilon^2)$;
☐ d. $O(\log(1/\epsilon))$.

20. Stochastic gradient descent is a descent method.
☐ True
☐ False