Quiz_ Quiz 6_...



# Quiz 6: Optimization

Started: Oct 21 at 1:39am

## Quiz Instructions

### Question 1 — 1 pts

Which of the following functions are **not** c-Lipschitz?

- $L(x) = \max(x, 0)$
- $L(x) = |x|$
- $L(x) = \sin(x)$
- $L(x) = \sqrt{x}$

### Question 2 — 1 pts





Which of the following method converges the fastest on any objective function?

- It is not possible to determine
- Newton's Method
- Adam
- AdaGrad

### Question 3 — 1 pts

Which of the following is **not** true about the convergence of the Gradient Descent method?

- The gap between the minimal objective value and the objective value at iteration $t$ is at most $\Theta(\frac{1}{t})$.
- If the objective function is convex and c -Lipschitz, gradient descent with a sufficiently small step size is guaranteed to converge to the minimal objective value.
- The gap between the minimal objective value and the objective value at iteration $t$ is at most $\Theta(\frac{1}{\sqrt{t}})$.
- To get to a parameter vector whose objective value that is $\epsilon$ larger than the minimal objective value, need $\Theta(\frac{1}{\epsilon^2})$ iterations.





### Question 4 — 1 pts

Which of the following statement refers to **Mini-batch Gradient Descent** method?

- Its smallest number of iterations needed to perform one pass over the dataset is 1.
- It computes the gradient on the entire dataset every iteration.
- It computes the gradient on a single data point every iteration.
- It computes the gradient on a subset of data points every iteration.





### Question 5 — 1 pts

Which is NOT true about the benefits of using Stochastic Gradient Descent (SGD), compared to vanilla Gradient Descent?

- The convergence path of SGD is less noisier than that of vanilla gradient descent.
- Its computation is cheaper.
- The convergence of SGD is much faster.
- SGD is less likely to stuck at local minimum or saddle point.

### Question 6 — 1 pts

Which is NOT true about tuning hyperparameters in iterative optimization algorithms?

- If the parameter vector is low-dimensional and we are using a stochastic method, try a inverse time learning rate schedule.





- If object value doesn't converge, try a larger mini-batch size.
- If objective value keeps the increasing trend without going down, trying a larger learning rate.
- It is a good idea to try different learning rate values uniformly sampled from a logarithmic scale.

### Question 7 — 1 pts

When using a model that's linear in the parameters, which is usually the best method for training?

- Stochastic gradient descent
- AdaGrad
- Newton's method
- Solving the normal equations

### Question 8 — 1 pts





What is the optimization constraint when minimizing a square loss function involving a multilayer perceptron model?

- There are no constraints
- The parameters are bounded by a maximum 2-norm
- The parameters are non-zero
- The parameters are bounded by a maximum 1-norm

No new data to save. Last checked at 1:39am Submit Quiz