Regression and Optimization

Email *

raphelwei@gmail.com

Your Name *

Sihan Wei

If you update the gradient of the loss function using one example, it's called * 1 point

- (●) Stochastic gradient descent
- ( ) Mini-batch gradient descent
- ( ) Batch gradient descent
- ( ) Other:

What does the regularization parameter in Ridge Regression control? * 1 point

- ○ The complexity of the model by setting a minimum threshold for feature selection.
- ◉ The trade-off between fitting the training data well and keeping the model coefficients small.
- ○ The number of iterations required for the model to converge.
- ○ The degree of polynomial features in the regression model.

If you update the gradient of the loss function using all examples, it's called * 1 point

- ○ Stochastic gradient descent
- ○ Mini-batch gradient descent
- ◉ Batch gradient descent
- ○ Other:

If my model has low loss, i.e. it's parameters are close to the global optimum of a convex objective, does that guarantee that I have low test loss? * 1 point

- ◉ No
- ○ Yes
- ○ I don't know.
- ○ Other:

Linear regression models naturally provide explanations for predictions because * 1 point

- ( ) They are linear models
- (x) The feature coefficients indicate the relative importance of each feature
- ( ) Squared error is similar to accuracy

What reasons would lead us to use L1 regularization instead of L2? (Select all) * 1 point

- [ ] L1 is easier to optimize
- [x] If we believe the optimal solution should be sparse
- [x] If we want the resulting model to be small (use as few features as possible)
- [ ] We want to use stochastic gradient descent

Which statement is true about convex functions? (select all) * 1 point

- [x] If x is a local minimum of a convex function f, then x is also the global minimum of f.
- [x] Let f be convex and differentiable, if f'(x)=0 (critical point), the x is a global minimum.
- [ ] A cubic function f(x)=x^3 is convex.
- [ ] A convex function can have only one global minimizer.

Which statement is true about the relationship between the learning rate and convergence in gradient descent? * 1 point

- ( ) A larger learning rate always speeds up convergence.
- ( ) A smaller learning rate guarantees convergence to the global minimum.
- (•) Too large of a learning rate may cause the algorithm to overshoot the minimum.
- ( ) The learning rate has no impact on convergence

What is the typical effect of choosing a very small learning rate for gradient descent optimization? * 1 point

- ( ) The optimization will converge to the global minimum very quickly.
- (•) The optimization may converge very slowly, increasing computation time.
- ( ) The optimization will become unstable, with large oscillations in the loss.
- ( ) The optimization will not be affected by the learning rate.

In the context of stochastic gradient descent (SGD), what does the term 'stochastic' refer to? * 1 point

- (•) The randomness in choosing the data points for each update.
- ( ) The random initialization of the model's parameters.
- ( ) The probability distribution of the model's predictions.
- ( ) The unpredictable nature of the optimization process.

Google Forms