Programming Assignment 5

Click this link to download the Ridge Regression notebook and then complete problems 1-3.

If you are completely stuck on this notebook, you can consult these hints.

Problem 1

1/1 point (graded)

In setting the step size, does it make sense in practice to pick a fixed schedule like $1/t$ (where $t = 1, 2, 3, \ldots$ is the iteration)? Select all that apply.

- [ ] Yes, that should work well in general.
- [x] Not really: it is important to adapt to the scale of the loss function and/or gradient; otherwise, the steps might be much too big or much too small.
- [ ] Not really: there is no need to reduce the step size with time.
- [ ] Yes: in fact any schedule that decreases with t will work well.

Submit

Problem 2

1/1 point (graded)

Is it possible to design the gradient descent scheme so that the loss values are monotonically decreasing with successive iterations?

- ○ Yes: in fact, this is guaranteed to happen in any case.
- ○ No: we cannot hope for this if we want guaranteed convergence.
- ◉ Yes: this can be ensured by choosing a step size adaptively.
- ○ No: in general it is not possible to avoid having the loss function go up and down.

✔

Submit

Problem 3

1/1 point (graded)

In the notebook, the data was generated using a *sparse* regression function w, in which only 10 of the 100 features were set. Did your solver do a good job of identifying the relevant features? Select all that apply.

- ☑ Yes, in the sense that the 10 highest-valued coordinates were the relevant features.
- ☑ In general, we should not expect ridge regression to yield sparse solutions.
- ☐ Yes: the coefficients for the relevant features were at least an order of magnitude larger than the coefficients for other features.

✔

Submit