Summer Python '23
Gradient Descent Exercise

Andrew Papanicolaou

August 8, 2023

#1.) Generate 1000 IID normals with mean $\mu = 1$ and variance $\sigma^2 = 4$.

#2.) Use gradient descent to find the maximum likelihood estimator (MLE) for this data under the assumption that we do not know (μ, σ). That is, find

$$(\hat{\mu}, \hat{\sigma}) = \arg\max_{\mu,\sigma} f(x_1, x_2, \ldots, x_n | \mu, \sigma) \ ,$$

where $x_i \sim iid\ normal(\mu, \sigma^2)$ and

$$\ln f(x_1, x_2, \ldots, x_n | \mu, \sigma) = -n \ln(\sigma\sqrt{2\pi}) - \frac{1}{2} \sum_{i=1}^{n} \frac{(x_i - \mu)^2}{\sigma^2} \ .$$

Use `learning_rate=.0001` and run for 1000 epoch. The gradient of f w.r.t. μ and σ is calculable explicitly.

#3.) How do your results compare with the estimated values obtained from the explicit formula for $(\hat{\mu}, \hat{\sigma})$?

#4.) How do the results change if you use random mini-batch with `batch_size=10`?