Optimization

Computational Mathematics and Statistics Camp
University of Chicago
September 2018

1. Find the gradient ∇f of the following functions and evaluate them at the given points.
   a. $f(x, y) = \sqrt{x^2 + y^2}, \quad (x, y) = (3, 4)$
   b. $f(x, y, z) = (x + z)e^{x-y}, \quad (x, y, z) = (1, 1, 1)$
2. Find the Hessian H for the following functions.
   a. $g(x, y) = x^4 - 3x^2y^3$
   b. $f(x, y, z) = xyz - x^2$
3. Find the local minimum values, local maximum values, and saddle point(s) of the function. Remember the process we discussed in class: Calculate the gradient, set it equal to zero to solve the system of equations, calculate the Hessian, and assess the Hessian at critical values. Be sure to show your work on each of these steps.
   a. $f(x, y) = x^4 + y^4 - 4xy + 2$
   b. $k(x, y) = (1 + xy)(x + y)$
4. Suppose we were interested in learning about how years of schooling affect the probability that a person turns out to vote. To simplify things, let's say we're trying to predict whether one individual voted. We did some preliminary work on this yesterday, but suppose we showed a colleague our model from this problem and they complained. "What a lame model", our colleague said, "You definitely have to include an intercept term." So in this problem we'll follow our colleague's advice and do just that. Let Y be our single observation of the dependent variable (whether or not a person turned out to vote) and X_1 be our single observation of an independent variable, *education*, the number of years of schooling for this individual. Now though, we're also going to include an intercept term, β_0 in our model along with β_1 a coefficient that's associated with X_1. This produces the following model for which we want to find the values of both β_0 and β_1 that minimize the sum of square errors.

   $$Y = \beta_0 + \beta_1 X_1 + \epsilon$$

   where ϵ is an error term. Use the method of least squares to solve for the values of β_0, β_1 that minimizes the sum of squared errors in the our data. Using the tools of multivariate minimization we've been practicing, find the values of β_0 and β_1 that minimize this quantity.
5. Suppose we were interested in learning about how years of schooling affect the probability that a person turns out to vote. To simplify things, let's say we're trying to predict whether one individual voted. Unlike question 4, let's assume we now have many observations in our data. Let Y_i be a vector containing many (n) observations of our dependent variable (whether or not a person i turned out to vote) and X_i be a vector containing many (n) observations of our independent variable, *education*, the number of years of schooling for individual i. We're also going to include an intercept term, β_0 in our model along with β_1 a coefficient that's associated with X_i. This produces the following model for which we want to find the values of both β_0 and β_1 that minimize the sum of square errors.

   $$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$$

   where ϵ_i is an error term for person i.
   a. Use the method of least squares to solve for the values of β_0, β_1 that minimizes the sum of squared errors in the our data. Using the tools of multivariate minimization we've been practicing, set the partial derivatives of β_0 and β_1 equal to 0 and determine $\hat{\beta}_0$ and $\hat{\beta}_1$.

   $$\sum_{i=1}^{n} (\epsilon_i)^2 = \sum_{i=1}^{n} (Y_i - (\beta_0 + \beta_1 X_i))^2$$

   b. **OPTIONAL:** Use what we learned about multivariate optimization to find the the Hessian of $\hat{\beta}$. Does this hessian indicate that $\hat{\beta}_0$ and $\hat{\beta}_1$ minimize the sum of squared errors?

- We've worked a lot on taking derivatives, but doing so with when summation is involved makes this different than what we've done before. Algebra tips for working with summation signs are below.
  - If you sum a constant n times, this is equivalent to multiplying that constant by n (e.g., $\sum_{i=1}^{n} 1 = n \times 1 = n$).
  - If you sum a variable that is also multiplied by a constant, you can move the constant outside the summation sign (e.g., $\sum_{i=1}^{n} 2 \times X_i = 2 \sum_{i=1}^{n} X_i$).
  - The summation sign is a linear operator so we can distribute it across variables we add together (e.g., $\sum_{i=1}^{n}(X_i + Y_i) = \sum_{i=1}^{n} X_i + \sum_{i=1}^{n} Y_i$)