1. Linear Approximation

A common way to approximate a nonlinear function is to perform linearization near a point. In the case of a one-dimensional function \( f(x) \), the linear approximation of \( f(x) \) at a point \( x_\star \) is given by

\[
\tilde{f}(x; x_\star) = f(x_\star) + f'(x_\star) \cdot (x - x_\star),
\]

(1)

where \( f'(x_\star) := \frac{df}{dx}(x_\star) \) is the derivative of \( f(x) \) at \( x = x_\star \).

Keep in mind that wherever we see \( x_\star \), this denotes a constant value or operating point.

We can evaluate the accuracy of our approximation by calculating the approximation error, namely \( |f(x) - \tilde{f}(x; x_\star)| \).

Suppose we have the single-variable function \( f(x) = x^3 - 3x^2 \). We can plot the function \( f(x) \) as follows:

![Plot of \( f(x) = x^3 - 3x^2 \)](image)

(a) Write the linear approximation of the function around an arbitrary point \( x_\star \).
(b) Using the expression above, linearize the function around the point \( x_\star = 1.5 \). Draw the linearization into the plot in fig. 1. Then evaluate the accuracy of the linear approximation at \( x = 1.7 \) and \( x = 2.5 \). Does the difference in accuracy make sense, based on the plot?

Now, we can extend this to higher dimensional functions. In the case of a two-dimensional function \( f(x, y) \), the linear approximation of \( f(x, y) \) at a point \((x_\star, y_\star)\) is given by

\[
\tilde{f}(x, y; x_\star, y_\star) = f(x_\star, y_\star) + \frac{\partial f}{\partial x}(x_\star, y_\star) \cdot (x - x_\star) + \frac{\partial f}{\partial y}(x_\star, y_\star) \cdot (y - y_\star).
\]

(2)

where \( \frac{\partial f}{\partial x}(x_\star, y_\star) \) is the partial derivative of \( f(x, y) \) with respect to \( x \) at the point \((x_\star, y_\star)\), and similarly for \( \frac{\partial f}{\partial y}(x_\star, y_\star) \).

(c) Now, let’s see how we can find partial derivatives. When we are given a function \( f(x, y) \), we calculate the partial derivative of \( f \) with respect to \( x \) by fixing \( y \) and taking the derivative with respect to \( x \). Given the function \( f(x, y) = x^2y \), find the partial derivatives \( \frac{\partial f(x, y)}{\partial x} \) and \( \frac{\partial f(x, y)}{\partial y} \).

(d) Write out the linear approximation of \( f \) near \((x_\star, y_\star)\).

(e) We want to see if the approximation arising from linearization of this function is reasonable for a point close to our point of evaluation. Suppose we want to evaluate the accuracy of our approximation at some point \((x_\star + \delta, y_\star + \delta)\), where \( x_\star = 2 \) and \( y_\star = 3 \). Find the accuracy of this approximation in terms of \( \delta \). What if \( \delta = 0.01 \)?
(f) Suppose we have now a scalar-valued function \( f(\vec{x}, \vec{y}) \), which takes in vector-valued arguments \( \vec{x} \in \mathbb{R}^n, \vec{y} \in \mathbb{R}^k \) and outputs a scalar \( \in \mathbb{R} \). That is, \( f(\vec{x}, \vec{y}) \) is \( \mathbb{R}^n \times \mathbb{R}^k \to \mathbb{R} \).

One way to linearize the function \( f \) is to do it for every single element in \( \vec{x} = [x_1 \ x_2 \ \ldots \ x_n]^\top \) and \( \vec{y} = [y_1 \ y_2 \ \ldots \ y_k]^\top \). Then, when we are looking at \( x_i \) or \( y_j \), we fix everything else as constant. This would give us the linear approximation
\[
f(\vec{x}, \vec{y}) \approx f(\vec{x}_\star, \vec{y}_\star) + \sum_{i=1}^{n} \frac{\partial f(\vec{x}, \vec{y})}{\partial x_i} \bigg|_{(\vec{x}_\star, \vec{y}_\star)} (x_i - x_i) + \sum_{j=1}^{k} \frac{\partial f(\vec{x}, \vec{y})}{\partial y_j} \bigg|_{(\vec{x}_\star, \vec{y}_\star)} (y_j - y_j) \quad \text{(3)}
\]

In order to simplify this equation, we can define the following two vector quantities:
\[
J_{\vec{x}f} = \begin{bmatrix} \frac{\partial f}{\partial x_1} & \cdots & \frac{\partial f}{\partial x_n} \end{bmatrix}
\]
\[
J_{\vec{y}f} = \begin{bmatrix} \frac{\partial f}{\partial y_1} & \cdots & \frac{\partial f}{\partial y_k} \end{bmatrix}
\]

First, how can we “vectorize” eq. (3) using \( J_{\vec{x}f} \) and \( J_{\vec{y}f} \)? Next, assume that \( n = k \) and we define the function \( f(\vec{x}, \vec{y}) = \vec{x}^\top \vec{y} = \sum_{i=1}^{k} x_i y_i \). Find \( J_{\vec{x}f} \) and \( J_{\vec{y}f} \) for this specific \( f \).

(HINT: For vectorizing, think about replacing the summations as the multiplication of a row and column vector. What would these vectors be?)

(g) Following the above part, find the linear approximation of \( f(\vec{x}, \vec{y}) \) near \( \vec{x}_\star = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \) and \( \vec{y}_\star = \begin{bmatrix} -1 \\ 2 \end{bmatrix} \). Recall that \( f(\vec{x}, \vec{y}) = \vec{x}^\top \vec{y} = \sum_{i=1}^{k} x_i y_i \).
These linearizations are important for us because we can do many easy computations using linear functions.

Contributors:

- Neelesh Ramachandran.
- Kuan-Yun Lee.