Tutorial Business Analytics

Exercise 12

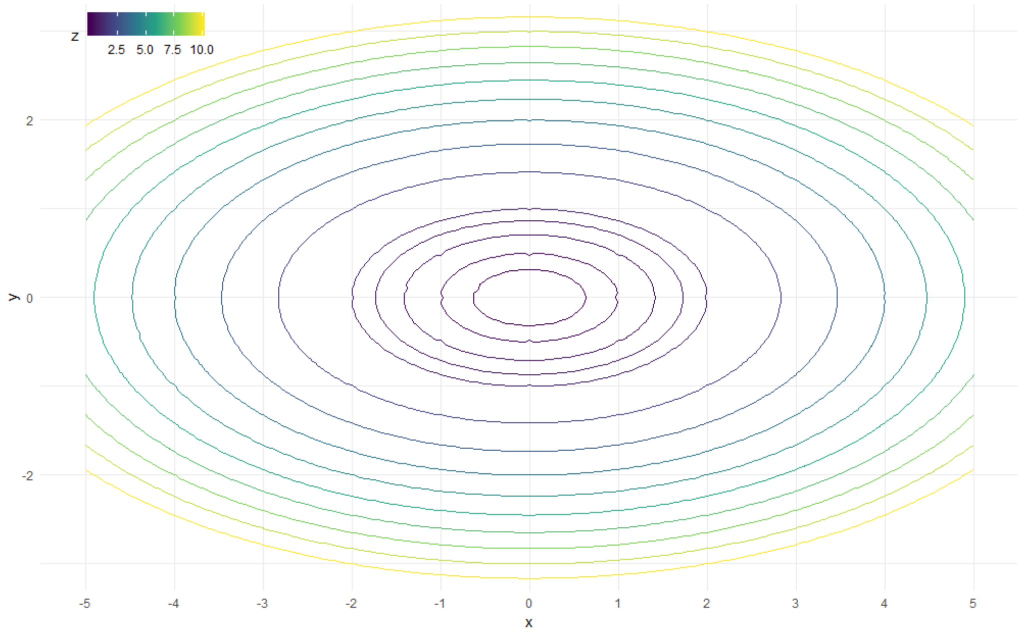
Exercise 12.1 Gradient Descent

Let $f: \mathbb{R}^2 \to \mathbb{R}$ be the convex function given by

$$f(x, y) = \frac{x^2}{4} + y^2$$

a) Find $\nabla f(x, y)$.

b) Starting from $(x_0, y_0) = (3, 2)^T$ perform 3 steps of gradient descent with a learning rate of $\alpha = 1$. Plot your gradient steps. What do you observe? Does the function value decrease in each step? Will this sequence converge to the optimum at $(0,0)$?



c) Repeat c) but with a learning rate rule that is guaranteed to converge:

$$\alpha_n = \frac{1}{n}$$

d) Starting from (x_1, y_1) found in b), perform 2 additional steps of the *momentum method*, with $\beta = 0.25$ and $\alpha = 1$. Assume that $d_1 = \nabla f(x_0, y_0)$

Recall from tutorial slides:

$$d_n = \beta d_{n-1} + \alpha \nabla f(x_{n-1}), \qquad x_n = x_{n-1} - d_n.$$

Exercise 12.2 Backpropagation I

Consider the following feed-forward neural network that consists of

- An input layer ($l = 0$) representing two-dimensional points

$$a^{[0]} = \left(a_1^{[0]}, a_2^{[0]}\right)^T \in \mathbb{R}^2$$

- A hidden layer $l = 1$ with 2 hidden nodes and sigmoid activation function $g^{[1]}$
- An output layer $l = 2$ with one node and sigmoid activation function $g^{[2]}$.

a) Write down the formulas for the forward pass of this neural network.
   How many trainable parameters does it have?

b) In the following, we will train the NN for binary classification on a data set. For a given input-output pair (x, y), we will use x as the input to the NN, model $y \approx \hat{y} = a^{[2]}(x)$ and evaluate the model using the *cross-entropy loss*

$$\ell(y, \hat{y}) = -[y \ \ln \hat{y} \ + (1 - y) \ \ln(1 - \hat{y})]$$

   Calculate the partial derivatives $\frac{\partial \ell}{\partial W_1^{[2]}}$ and $\frac{\partial \ell}{\partial W_2^{[2]}}$ that will be used to update $W^{[2]}$ in backpropagation.

   Hint: You may use the following derivative of the sigmoid function $\sigma(\cdot)$ without proof: $\sigma'(x) = \sigma(x) \cdot (1 - \sigma(x))$.