CAS CS 365
Lab 10

1. Chain rule
   a. dz/dt of the following functions
   b.
      1. $z = f(x, y) = 4x^2 + 3y^2,\ x = x(t) = \sin t,\ y = y(t) = \cos t$
      2. $z = f(x, y) = \sqrt{x^2 - y^2},\ x = x(t) = e^{2t},\ y = y(t) = e^{-t}$

      i.
         1. $2 \sin t \, \cos t$
         2. $\dfrac{2e^{6t} + 1}{e^t \sqrt{e^{6t} - 1}}$
2. Chain rule for two variables
   a. Calculate partial derivatives using the following functions
   b. $z = f(x, y) = 3x^2 - 2xy + y^2, x = x(u, v) = 3u + 2v, y = y(u, v) = 4u - v$

      i.
         $$\frac{\partial z}{\partial u} = 38u + 18v$$
         $$\frac{\partial z}{\partial v} = 18u + 34v$$
3. Gradient (partial derivative)
   a. Direction of the greatest change of a function (value is the same as partial derivative)
   b. Find the gradient of the following function
   c. $f(x, y) = x^2 - xy + 3y^2$

      i. $\nabla f(x, y) = [2x - y, -x + 6y]$
4. More about Gradient
   a. Find the gradient of the following
   b. $f(x, y, z) = e^{-2z} \sin 2x \cos 2y$

      i. $\nabla f(x, y, z) = 2e^{-2z} \cdot [\cos 2x \cos 2y, -\sin 2x \sin 2y, -\sin 2x \cos 2y]$
5. Directional derivative
   a. Let theta = arccos(⅗), find the directional derivative of

      $f(x, y) = x^2 - xy + 3y^2$ In the direction of $v = (\cos\theta,\ \sin\theta)$

      i.
         Partial derivative of f is $[2x - y, -x + 6y]$

         $$\nabla_v f(x, y) = (2x - y)\frac{3}{5} + (-x + 6y)\frac{4}{5} = \frac{2x + 21y}{5}$$

6. Gradient of a least-square loss in a linear model
    a. Consider the linear model y = X * theta where theta is a parameter vector of length D, X is an n by D input feature matrix and y are the the corresponding observations of length n
    b. Optimizing such a model can be considered as solving $\min_{\theta \in \mathbb{R}^{\mathbb{D}}} \left(||y - X \cdot \theta||^2 \right)$
    c. This can be solved by computing the gradient of $L = ||e||^2, \; e = y - X \cdot \theta$

$$\frac{\partial L}{\partial e} = 2e^T \qquad \frac{\partial e}{\partial \theta} = -X$$

    d. $$\frac{\partial L}{\partial \theta} = \frac{\partial L}{\partial e} \frac{\partial e}{\partial \theta} = -2\left(y^T - \theta^T X^T\right) X$$
    e. Solving derivative equals 0 is sufficient to minimize the loss since the Hassian of L equals X^T*X is PSD