CS 189: Introduction to Machine Learning - Discussion 6

1. Fun with Lagrange Multipliers (taken from Harvard class worksheet)

   (a) Minimize the function:
   $f(x, y, z) = x + y + 2z$ such that $x^2 + y^2 + z^2 = 3$

   (b) Minimize the function:
   $f(x, y, z) = x^2 - y^2$ such that $x^2 + 2y^2 + 3z^2 = 1$

Solution:

(a) The Lagrangian is:

$$L(x, y, z, \lambda) = x + y + 2z + \lambda(x^2 + y^2 + z^2 - 3)$$

Taking all of the partial derivatives and setting them to 0, we get this system of equations:

$$\begin{aligned}\lambda x &= -\frac{1}{2}\\ \lambda y &= -\frac{1}{2}\\ \lambda z &= -1\\ x^2 + y^2 + z^2 &= 3\end{aligned}$$

We can infer that $x = y$, and that $z = 2y$. Plugging this into the constraint, we have:

$$y^2 + y^2 + 4y^2 = 3$$

which shows that $y = \pm\frac{\sqrt{2}}{2}$. We have two critical points, $(-\frac{\sqrt{2}}{2}, -\frac{\sqrt{2}}{2}, -\sqrt{2})$ and $(\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}, \sqrt{2})$. Plugging these into our objective function f, we find that the minimizer is the former, with a value of $-3\sqrt{2}$.

(b) The Lagrangian is:

$$x^2 - y^2 + \lambda(x^2 + 2y^2 + 3z^2 - 1)$$

Taking all of the partial derivatives and setting them to 0, we get this system of equations:

$$\begin{aligned}x &= -\lambda x\\ y &= 2\lambda y\\ 0 &= \lambda z\\ x^2 + 2y^2 + 3z^2 &= 1\end{aligned}$$

To solve this, we look at several cases:

Case 1: $\lambda = 0$. This implies that $x = y = 0$, and $z = \pm\frac{1}{3}$. We have two critical points: $(0, 0, \pm\frac{1}{3})$.

Case 2: $\lambda \neq 0$. z must be 0.

Case 2a: $x = 0$. The constraint gives us that $y = \pm\frac{1}{\sqrt{2}}$. This gives us another two critical points: $(0, \pm\frac{1}{\sqrt{2}}, 0)$.

Case 2b: $y = 0$. The constraint gives us $x = \pm 1$, giving us another two critical points: $(\pm 1, 0, 0)$.

Plugging in all of our critical points, we find that $(0, \pm\frac{1}{\sqrt{2}}, 0)$ minimizes our function with a value of $-\frac{1}{2}$.

2. Quadratic Kernel

   Find a feature mapping Φ such that $\Phi(x)^T\Phi(y) = K(x, y)$ where the kernel function is $K(x, y) = (x^T y + 1)^2$. For simplicity, you may assume that the data is 2-dimensional, i.e. $x = [x_1, x_2]^T$.

Solution:

$$\begin{aligned}
K(x, y) = (x^T y + 1)^2 &= (x_1 y_1 + x_2 y_2 + 1)^2 \\
&= 1 + x_1^2 y_1^2 + x_2^2 y_2^2 + 2x_1 y_1 + 2x_2 y_2 + 2x_1 y_1 x_2 y_2 \\
&= 1 + x_1^2 y_1^2 + x_2^2 y_2^2 + 2x_1 y_1 + 2x_2 y_2 + 2x_1 y_1 x_2 y_2 \\
&= [1, \sqrt{2}x_1, \sqrt{2}x_2, \sqrt{2}x_1 x_2, x_1^2, x_2^2][1, \sqrt{2}y_1, \sqrt{2}y_2 \sqrt{2}y_1 y_2, y_1^2, y_2^2]^T \\
&= \Phi(x)^T \Phi(y)
\end{aligned}$$

where

$$\Phi(x) = [1, \sqrt{2}x_1, \sqrt{2}x_2, \sqrt{2}x_1 x_2, x_1^2, x_2^2]^T$$

3. Fun with Newton's method for root-finding

(a) Write down the iterative update equation of Newton's method for finding a root $x : f(x) = 0$ for a real-valued function f.

Solution: $x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}$

(b) Prove that if $f(x)$ is a quadratic function ($f(x) = ax^2 + bx + c$), then it only takes one iteration of Newton's Method to find the minimum/maximum.

Solution: The Newton's method update for finding a mininum/maximum is

$$x_{n+1} = x_n - \frac{f'(x_n)}{f''(x_n)} = x_n - \frac{2ax_n + b}{2a} = \frac{-b}{2a}$$

And this is the point for mininum/maximum.