STA4273_HW3

Alex Jones

2023-04-09

Question 1

Question 1 asks to determine whether the denoted function is concave, convex, or neither.

(a) With $f(x) = e^x - 1$, we calculate the gradient to be $\nabla f(x) = e^x$ and the Hessian to be $\nabla^2 f(x) = e^x > 0$ for all $x \in \mathbb{R}$, indicating the function is convex.

(b) With $f(x) = x_1 x_2$, we calculate the gradient to be $\nabla f(x) = \begin{bmatrix} x_2 \\ x_1 \end{bmatrix}$ and the Hessian to be $\nabla^2 f(x) = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$, which is neither positive or negative semi-definite. As such, it is neither convex or concave.

(c) With $f(x) = x_1^\alpha x_2^{1-\alpha}$, we find the gradient to be $\nabla f(x) = \begin{bmatrix} \alpha x_1^{\alpha-1} x_2^{1-\alpha} \\ (1-\alpha) x_1^\alpha x_2^{-\alpha} \end{bmatrix}$ and the Hessian to be $\nabla^2 f(x) = \begin{bmatrix} \alpha(\alpha-1) x_1^{\alpha-2} x_2^{1-\alpha} & \alpha(1-\alpha) x_1^{\alpha-1} x_2^{-\alpha} \\ \alpha(1-\alpha) x_1^{\alpha-1} x_2^{-\alpha} & -\alpha(1-\alpha) x_1^\alpha x_2^{-\alpha-1} \end{bmatrix}$ Which is negative semi-definite. As such, we find $f(x)$ to be concave.

Question 2

Question 2 asks to prove the concavity of the "logsumexp" function.

(a) With $f(x) = log(e^{x_1} + e^{x_2})$, we find the gradient to be
$\nabla f(x) = \begin{bmatrix} \frac{e^{x_1}}{e^{x_1}+e^{x_2}} \\ \frac{e^{x_2}}{e^{x_1}+e^{x_2}} \end{bmatrix}$
and the Hessian to be
$\nabla^2 f(x) = \begin{bmatrix} \frac{e^{x_1+x_2}}{(e^{x_1}+e^{x_2})^2} & \frac{-e^{x_1+x_2}}{(e^{x_1}+e^{x_2})^2} \\ \frac{-e^{x_1+x_2}}{(e^{x_1}+e^{x_2})^2} & \frac{e^{x_1+x_2}}{(e^{x_1}+e^{x_2})^2} \end{bmatrix}$
$= \frac{e^{x_1+x_2}}{(e^{x_1}+e^{x_2})^2} \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$
The matrix is positive semi-definite, and thus the function is convex when $p = 2$

(b) By restricting to a line, let $g(t) = log(\sum_{i=1}^n exp(\vec{z_i} + t\vec{v_i}))$
Then $g'(t) = \frac{\sum_{i=1}^n \vec{v}_i exp(\vec{z}_i + t\vec{v}_i)}{\sum_{i=1}^n exp(\vec{z}_i + t\vec{v}_i)}$
and
$g''(t) = \frac{(\sum_{i=1}^n exp(\vec{z}_i + t\vec{v}_i))(\sum_{i=1}^n \vec{v}_i^2 exp(\vec{z}_i + t\vec{v}_i)) - (\sum_{i=1}^n \vec{v}_i exp(\vec{z}_i + t\vec{v}_i))^2}{\sum_{i=1}^n exp(\vec{z}_i + t\vec{v}_i)^2}$
Since the denominator is always positive, we must prove that numerator is ≥ 0. Using the Cauchy-Schwarz inequality, which in the context of this problem states:
$\sum_{i=1}^n exp(x_i) \sum_{i=1}^n v_i^2 exp(x_i) \geq (\sum_{i=1}^n (v_i exp(x_i)))^2$
We can see the numerator is always positive.
Because of this, we can conclude that $g(t)$ is convex, and as such the logsumexp function is convex.

Question 3

Question 3 asks to create and implement the gradient descent method on the below function.

(a) Rewriting the original function as $f(x) = x_1^2 + x_1 x_2 + x_2^2 + x_2 x_3 + x_3^2 + \sum_{i=1}^{3} e^{x_i}$, we calculate the gradient to be:

$$\nabla f(x) = \begin{bmatrix} 2x_1 + x_2 + e^{x_1} \\ x_1 + 2x_2 + x_3 + e^{x_2} \\ x_2 + 2x_3 + e^{x_3} \end{bmatrix}$$

(b) Implementing the algorithm:

```r
swag <- function(x) {
  x[1]**2+x[1]*x[2]+x[2]**2+x[2]*x[3]+x[3]**2+exp(x[1])+exp(x[2])+exp(x[3])
}

swagGrad <- function(x) {
  c(2*x[1]+x[2]+exp(x[1]), x[1]+2*x[2]+x[3]+exp(x[2]), x[2]+2*x[3]+exp(x[3]))
}

step_size = 0.01
x <- c(1,1,1)

for (i in 1:1000) {
  x <- x - step_size*swagGrad(x)
}
```

We find that the minimum value of the function is 2.6345649 with the values of x_1, x_2, and x_3 being -0.3022136, -0.1347526, -0.3022136, respectively.