Worksheet on "Background on calculus/optimization, Density Estimation"

PRML – CS5691 (Jul–Nov 2023)

September 11, 2023

1. a Find the linear approximation of $f(x) = \sqrt{x}$ at $x = 16$

   b Use it to approximate $\sqrt{15.9}$

Solution:
Part a:
$f'(x) = \dfrac{d}{dx}\sqrt{x} = \dfrac{1}{2\sqrt{x}}$
$f'(16) = \dfrac{1}{2\sqrt{16}} = \dfrac{1}{8}$
$L(x) = f'(a)(x-a) + f(a) = \dfrac{1}{8}(x-16) + \sqrt{16} = \dfrac{x}{8} + 2$
Part b:
$\sqrt{15.9} = f(15.9) \approx \frac{1}{8} \cdot 15.9 + 2 = \frac{319}{80}$
Source link

2. Find the tangent plane to $f(x, y) = 2 - x^2 - y^2$ at $(\frac{1}{2}, -\frac{1}{2})$

Solution: First we compute the partial derivatives at $(\frac{1}{2}, -\frac{1}{2})$
$\frac{\partial f}{\partial x} = -2x = -1$ and
$\frac{\partial f}{\partial y} = -2y = 1$
Since, $f(1/2, -1/2) = 3/2$, we see from the theorm below (Figure 1),

$$f(x, y) \approx L(x, y) = f(x_0, y_0) + f_x(x_0, y_0)(x - x_0) + f_y(x_0, y_0)(y - y_0)$$

Figure 1

that the equation of tangent plane is $z = 3/2 - (x - 1/2) + (y + 1/2) = 5/2 - x + y$
Source link

3. Prove if the statement if true [or] Provide counter-example if the statement is false:

   a Sum of two convex functions is a convex function

   b Product of two convex functions is a convex function

   c Difference of two convex functions is a convex function

Solution:
Part a: True, Proof:
$f_1(c.x_1 \leq +(1-c).x_2)) \leq c.f_1(x_1) + (1-c).f_1(x_2)$ and
$f_2(c.x_1 \leq +(1-c).x_2)) \leq c.f_2(x_1) + (1-c).f_2(x_2)$
$f_1(c.x_1 \leq +(1-c).x_2)) + f_2(c.x_1 \leq +(1-c).x_2)) \leq c.f_1(x_1) + (1-c).f_1(x_2) + c.f_2(x_1) + (1-c).f_2(x_2)$
$f_1(c.x_1 \leq +(1-c).x_2)) + f_2(c.x_1 \leq +(1-c).x_2)) \leq c(f_1(x_1) + f_2(x_1)) + (1-c)(f_1(x_2) + f_2(x_2))$
$f(c.x_1 + (1-c).x_2) \leq c.f(x_1) + (1-c).f(x_2)$ where $f = f_1 + f_2$

Part b False, counterexample:
Functions $f(x) = 1+x$ and $g(x) = 1-x$ are convex fucntions, however their product $(f*g)(x) = 1-x^2$ is not a convex funciton. Ref

Part c: False, Counterexample:
Functions $f(x) = \sqrt{x^2+1}$ and $g(x) = |x|$ are convex, however their difference $(f-g)(x) = \sqrt{x^2+1} - |x|$ is not a convex function. Ref

Use GeoGebra for graph visualizations.

4. Suppose that a particular gene occurs as one of two alleles (A and a), where allele A has frequency θ in the population. That is, a random copy of the gene is A with probability θ and a with probability $1-\theta$. Since a diploid genotype consists of two genes, the probability of each genotype is given by:

Genotype	Probability
AA	θ^2
Aa	$2\theta(1-\theta)$
aa	$(1-\theta)^2$

Suppose we test a random sample of people and find that k1 are AA, k2 are Aa, and k3 are aa. Find the MLE of θ.

Solution:
The likelihood function is given by:

$$\mathcal{L}(\theta|k_1,k_2,k_3) = \theta^{2k_1} \cdot (2\theta(1-\theta))^{k_2} \cdot (1-\theta)^{2k_3}$$

So, the log-likelihood is given by:

$$\ln \mathcal{L}(\theta|k_1,k_2,k_3) = 2k_1 \ln(\theta) + k_2 \ln(\theta) + k_2 \ln(1-\theta) + 2k_3 \ln(1-\theta)$$

Now, set the derivative equal to zero:

$$\frac{d}{d\theta} \ln \mathcal{L}(\theta|k_1,k_2,k_3) = \frac{2k_1+k_2}{\theta} - \frac{k_2+2k_3}{1-\theta} = 0$$

Solving for θ, we find the maximum likelihood estimate (MLE) is:

$$\hat{\theta} = \frac{2k_1+k_2}{2(k_1+k_2+k_3)}$$

Reference

5. a Complete the derivation of MLE of Bernoulli Distribution seen in class
   b Similarly complete the derviation of MLE of Multinoulli Distribution.

   Hint: You can use log likelihood LL seen in class and follow the below steps:
   i Compute the gradient of log likelihood LL
   ii Equate it to zero to find the stationary points
   iii Argue the stationary point is global maxima - e.g., by verifying if the LL is concave

Solution:

a Let m be the no. of 1s and n be the total no. of tosses
We can write the log likekihood LL as :

$$\mathcal{L}(\mu|x_1, x_2, \ldots, x_n) = m\ln(\mu) + (n-m)\ln(1-\mu)$$

Now, differentiating the log-likelihood w.r.t μ and setting it to 0 to find the MLE:

$$\frac{\partial \mathcal{L}}{\partial \mu} = \frac{m}{\mu} - \frac{n-m}{1-\mu} = 0$$

Solving for μ:

$$\begin{aligned}\frac{m}{\mu} - \frac{n-m}{1-\mu} &= 0\\ &\Rightarrow m - m\mu = n\mu - m\mu\\ &\Rightarrow \mu = \frac{m}{n}\end{aligned}$$

To prove that LL is concave for μ we find the second derivate as follows:

$$\begin{aligned}\frac{\partial^2 l}{\partial \mu^2} &= -\frac{m}{\mu^2} - \frac{n-m}{(1-\mu)^2}\\ &= -\left(\frac{m}{\mu^2} + \frac{n-m}{(1-\mu)^2}\right)\\ &< 0 \ \forall \mu \in \mathbb{R}\end{aligned}$$

Therfore, The stationary point is the global maxima

b The log likelihood LL of a Multinoulli distribution using a Lagrange multiplier to ensure that $\sum_k \mu_k = 1$ is as follows:

$$\sum_k m_k \ln \mu_k + \lambda\left(\sum_k \mu_k - 1\right)$$

Now, differentiating the log-likelihood w.r.t μ_k and setting it to 0 to find the MLE we get:

$$\frac{m_k}{\mu_k} + \lambda = 0$$

$$\mu_k = \frac{-m_k}{\lambda}$$

To solve for λ, we sum both sides and make use of our initial constraint:

$$\sum_{k=1}^{K} \mu_k = \frac{-\sum_{k=1}^{K} m_k}{\lambda}$$

$$1 = \frac{-\sum_{k=1}^{K} m_k}{\lambda}$$

$$\lambda = -N$$

Therefore,

$$\mu_k = \frac{m_k}{N}$$

To prove that LL is concave for μ we find the second derivate as follows:

$$\frac{\partial^2 L}{\partial \mu_k^2} = \frac{-m_k}{\mu_k^2}$$

$$\leq 0 \; \forall \mu_k \in \mathbb{R}$$

$$\frac{\partial^2 L}{\partial \mu_k \partial \mu_k'} = 0$$

Arranging them to a Hessian matrix, we get $H(x) = \begin{bmatrix} \leq 0 & 0 \\ 0 & \leq 0 \end{bmatrix}$
Hessian is negative semidefinite because all off-diagonal entries are zero and on-diagonal will be negative or zero. The eigen values of such a matrix will be negative or zero. Therfore, it is a concave function and thereby the stationary point is the global maxima
The derivation can also be solved using another approach Reference

6. Prove that $\frac{\partial}{\partial x}(x^T A\, x) = A^T x + Ax$ (or $2Ax$ if A is Symmetric)

(Hint: $x^T A\, x = \sum_{j=1}^{n} \sum_{i=1}^{n} a_{ij} x_i x_j$)

Solution:

$$
\begin{aligned}
\mathbf{x}^T\mathbf{A}\mathbf{x} &= \begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix} \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \\
&= \begin{bmatrix} (a_{11}x_1 + \cdots + a_{n1}x_n) & \cdots & (a_{1n}x_1 + \cdots + a_{nn}x_n) \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \\
&= \begin{bmatrix} \sum_{i=1}^{n} a_{i1}x_i & \cdots & \sum_{i=1}^{n} a_{in}x_i \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \\
&= x_1 \sum_{i=1}^{n} a_{i1}x_i + \cdots + x_n \sum_{i=1}^{n} a_{in}x_i \\
&= \sum_{j=1}^{n} x_j \sum_{i=1}^{n} a_{ij}x_i \\
&= \sum_{j=1}^{n} \sum_{i=1}^{n} a_{ij}x_i x_j
\end{aligned}
$$

Continuing from the above results,

$$
\frac{\partial \mathbf{x}^T\mathbf{A}\mathbf{x}}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial \mathbf{x}^T\mathbf{A}\mathbf{x}}{\partial x_1} \\ \vdots \\ \frac{\partial \mathbf{x}^T\mathbf{A}\mathbf{x}}{\partial x_n} \end{bmatrix}
$$

Consider the k^{th} row in the above vector:

$$
\begin{aligned}
\frac{\partial \mathbf{x}^T\mathbf{A}\mathbf{x}}{\partial x_k} &= \frac{\partial}{\partial x_k} \left(\sum_{j=1}^{n} \sum_{i=1}^{n} a_{ij}x_i x_j \right) \\
&= \frac{\partial}{\partial x_k} \left(x_1 \sum_{i=1}^{n} a_{i1}x_i + \cdots + x_k \sum_{i=1}^{n} a_{ik}x_i + \cdots + x_n \sum_{i=1}^{n} a_{in}x_i \right) \\
&= x_1 a_{k1} + \cdots + \left(\sum_{i=1}^{n} a_{ik}x_i + x_k a_{kk} \right) + \cdots + x_n a_{kn} \\
&= \sum_{j=1}^{n} a_{kj}x_j + \sum_{i=1}^{n} a_{ik}x_i \\
&= (k^{\text{th}} \text{ row of } \mathbf{A})\mathbf{x} + (\text{transpose of } k^{\text{th}} \text{ column of } \mathbf{A})\mathbf{x} \\
&= \left[(k^{\text{th}} \text{ row of } \mathbf{A}) + (\text{transpose of } k^{\text{th}} \text{ column of } \mathbf{A}) \right] \mathbf{x}
\end{aligned}
$$

Therefore,

$$\begin{aligned}
\frac{\partial \mathbf{x}^T \mathbf{A}\mathbf{x}}{\partial \mathbf{x}} &= \begin{bmatrix} [(1^{\text{st}} \text{ row of } \mathbf{A}) + (\text{transpose of } 1^{\text{st}} \text{ column of } \mathbf{A})]\mathbf{x} \\ \vdots \\ [(n^{\text{th}} \text{ row of } \mathbf{A}) + (\text{transpose of } n^{\text{th}} \text{ column of } \mathbf{A})]\mathbf{x} \end{bmatrix} \\
&= \begin{bmatrix} [(1^{\text{st}} \text{ row of } \mathbf{A}) + (\text{transpose of } 1^{\text{st}} \text{ column of } \mathbf{A})] \\ \vdots \\ [(n^{\text{th}} \text{ row of } \mathbf{A}) + (\text{transpose of } n^{\text{th}} \text{ column of } \mathbf{A})] \end{bmatrix} \mathbf{x} \\
&= \left(\begin{bmatrix} (1^{\text{st}} \text{ row of } \mathbf{A}) \\ \vdots \\ (n^{\text{th}} \text{ row of } \mathbf{A}) \end{bmatrix} + \begin{bmatrix} (\text{transpose of } 1^{\text{st}} \text{ column of } \mathbf{A}) \\ \vdots \\ (\text{transpose of } n^{\text{th}} \text{ column of } \mathbf{A}) \end{bmatrix} \right) \mathbf{x} \\
&= \left(\mathbf{A} + \mathbf{A}^T \right) \mathbf{x}
\end{aligned}$$

Reference

7. a Prove that $f(x) = x_1 \cdot x_2$ is not convex.
   b Prove that $f(x) = {x_1}^2 + {x_2}^2$ is not convex.

Hint: Use the property, f is convext iff $H(x)$ is positive semidefinite $\forall x \in \mathbb{R}^d$

Solution: Part a:
Given function is $f(x) = x_1 \cdot x_2$
Gradient of the function, $\nabla f(x) = \begin{bmatrix} x_2 \\ x_1 \end{bmatrix}$ and Hessian matrix is calculated as $H(x) = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$
Compute the Eigen values of the above matrix, $\lambda_1 = 1$ and $\lambda_2 = -1$. Hence $H(x)$ is not positive semidefinite. Hence, $f(x)$ is not convex.

Part b:
Given function, $f(x) = {x_1}^2 + {x_2}^2$
Gradient of the function, $\nabla f(x) = \begin{bmatrix} 2x_1 \\ 2x_2 \end{bmatrix}$ and Hessian matrix is calculated as $H(x) = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$
Compute the Eigen values of the above matrix, $\lambda_1 = 2$ and $\lambda_2 = 2$. Hence $H(x)$ is positive semidefinite. Hence, $f(x)$ is convex.
Reference