D. Arvind Sai A20553046

1. Problem.

1. verify the following identity.

$$(Q^{-1} + B^{T}P^{-1}B)^{-1} B^{T}P^{-1} = QB^{T}(BQB^{T} + P)^{-1}$$

so, the given identity is.

$$(Q^{-1} + B^{T}P^{-1}B)^{-1} B^{T}P^{-1} = QB^{T}(BQB^{T} + P)^{-1}$$

we can solve left side of the derivation first

$$(Q^{-1} + B^{T}P^{-1}B)^{-1} B^{T}P^{-1}$$

now multiply both sides by $(BQ^{T}B^{T} + P)$ on right

$$(Q^{-1} + B^{T}P^{-1}B)^{-1} B^{T}P^{-1}(BQ^{T}B^{T} + P) = QB^{T}$$

we got the derivation after multiplying on right.

now multiply both sides on the left by

$$(Q^{-1} + B^{T}P^{-1}B).$$

$$B^{T}P^{-1}(BQ^{T}B^{T} + P) = (Q^{-1} + B^{T}P^{-1}B)\,QB^{T}$$

Simplify the left side

$$B^{T}P^{-1}(BQ^{T}B^{T} + P) = Q^{-1}QB^{T} + B^{T}P^{-1}BQB^{T}$$

Simplify the right side QB^T

we have $B^T P^{-1} (BQ^T B^T + P)$ on right side

and $Q^{-1} Q B^T$ on side.

after simplification.

$$B^T P^{-1} (BQ^T B^T + P) = Q^{-1} Q B^T$$

$$B^T P^{-1} B Q^T B^T + B^T P^{-1} P = Q B^T$$

Since matrix multiplication is associative arrange terms as you follow:

$$B^T P^{-1} (BQ^T B^T) + B^T P^{-1} P = Q B^T$$

P^{-1} in the P is identity matrix because any matrix multiply by its inverse it is identity matrix

$$B^T P^{-1} (BQ^T B^T) + B^T = Q B^T$$

Subtract B^T on both side.

$$B^T P^{-1} (BQ^T B^T) = -B^T + Q B^T$$

To isolate $B^T P^{-1} (BQ^T B^T)$, multiply both sides on left by $(BQ^T B^T)^{-1}$

$$(BQ^T B^T)^{-1} B^T P^{-1} (BQ^T B^T) = (BQ^T B^T)(QB^T - B^T)$$

$\therefore$ left side simplifies to P^{-1}

$$P^{-1} = (BQ^T B^T)^{-1} (QB^T - B^T)$$

finally to isolate P, take the inverse of both sides.

$$P = [(BQ^T B^T)^{-1} (QB^T - B^T)]^{-1}$$

This completes the derivation of given identity.

Q. Verify the following Woodberg identity.

$$(A+BD^{-1}C)^{-1} = A^{-1} - A^{-1}B(D+CA^{-1}+B)^{-1} CA^{-1}$$

given identity is.

$$(A+BD^{-1}C)^{-1} = A^{-1} - A^{-1}B(D+CA^{-1}B)^{-1} CA^{-1}$$

Multiply both sides on the right $D+CA^{-1}B$

$$(A+BD^{-1}C)^{-1}(D+CA^{-1}B) = A^{-1} - A^{-1}B(D+CA^{-1}B)^{-1}$$

$$CA^{-1}(D+CA^{-1}B)$$

∴ now multiply both sides on the left by

$A+BD^{-1}C$

$$(A+BD^{-1}C)^{-1}(D+CA^{-1}B) = (A+BD^{-1}C)A^{-1} -$$

$$(A+BD^{-1}C)A^{-1}B(D+CA^{-1}B)^{-1}CA^{-1}(D+CA^{-1}B)$$

simplify the left side.

$$(D+CA^{-1}B) = (A+BD^{-1}C)A^{-1} - (A+BD^{-1}C)$$

$$A^{-1}B(D+CA^{-1}B)^{-1}CA^{-1}(D+CA^{-1}B)$$

cancel terms on both sides.

$$(D+CA^{-1}B) = (A+BD^{-1}C)A^{-1} - (A+BD^{-1}C)A^{-1}B$$

since it is identity matrix we can arrange terms

$$(D+CA^{-1}B) = A^{-1}(A+BD^{-1}C) - A^{-1}(A+BD^{-1}C)B$$

factor out A^{-1}

$$(D+CA^{-1}B) = A^{-1}(A+BD^{-1}C - (A+BD^{-1}C)B)$$

$$(D+CA^{-1}B) = A^{-1}(A - AB + BD^{-1}CB)$$

∴ now isolate $(D+CA^{-1}B)$ by multiplying both sides by A.

$$A(D+CA^{-1}B) = A - AB + D^{-1}BCB$$

∴ now isolate $(D+CA^{-1}B)$ by dividing both sides by A.

$$D+CA^{-1}B = I - AB + BD^{-1}CB$$

Subtract $I-AB$ from both side.

$$D+CA^{-1}B - (I-AB) = BD^{-1}CB$$

simplify

$$D - I + CA^{-1}B + AB = BD^{-1}CB$$

Rearrange terms

$$(A+BD^{-1}C)-I = B(A-I+(A^{-1}B)D^{-1}CB.$$

isolate the left side

$$(A+BD^{-1}C)-I = B(A-I+(A^{-1}B)D^{-1}CB$$

∴ finally add I to both sides.

$$(A+BD^{-1}C) = I+B(A-I+(A^{-1}B)D^{-1}CB.$$

finally after the verification the equation/ derivative is.

$$(A+BD^{-1}C)^{-1} = A^{-1}-A^{-1}B(D+CA^{-1}B)^{-1}CA^{-1}$$

Hence proved.

Problem 2:

1. Given $x = [x_1 : x_2 : x_3] \in R^3$ and $y = [y_1 : y_2] \in R^2$ where $y_1 = x_1^2 - x_2$ and $y_2 = x_3^2 + 3x_2$ compute $\partial y / \partial x$.

differentiation in Partial ways.

$$\frac{\partial y}{\partial x} = \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_2}{\partial x_2} & \cdots & \frac{\partial y_m}{\partial x_1} \\ \frac{\partial y_1}{\partial x_2} & \frac{\partial y_2}{\partial x_2} & \cdots & \frac{\partial y_m}{\partial x_2} \\ \vdots & & & \\ \frac{\partial y_1}{\partial x_n} & \frac{\partial y_2}{\partial x_n} & & \frac{\partial y_m}{\partial x_n} \end{bmatrix}$$

$$\frac{\partial y_1}{\partial x_1} = \frac{\partial}{x_1}(x_1^2 - x_2)$$

$$= 2x_1$$

$$\frac{\partial y_2}{\partial x_1} = \frac{\partial}{\partial x_1}(x_3^3 + 3x_2)$$

$$= 0$$

$$\frac{\partial y_1}{\partial x_2} = \frac{\partial}{x_2}(x_1^2 - x_2)$$

$$\frac{\partial y_2}{\partial x_2} = \frac{\partial}{\partial x_2}(x_3^3 + 3x_2)$$

$$\frac{\partial y_3}{\partial x_3} = \frac{\partial}{\partial x_3}(x_1^2 - x_2)$$

$$= 0$$

$$\frac{\partial y_3}{\partial x_3} = \frac{\partial}{x_3}(x_3^3 + 3x_2)$$

$$= 2x_3 + 0$$

$$= 2x_3$$

$$\frac{\partial y}{\partial x} = \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_2}{\partial x_1} \\ \frac{\partial y_1}{\partial x_2} & \frac{\partial y_2}{\partial x_2} \\ \frac{\partial y_1}{\partial x_3} & \frac{\partial y_2}{\partial x_3} \end{bmatrix}$$

$$\frac{\partial y}{\partial x} = \begin{bmatrix} 2x_1 & 0 \\ -1 & 6x_2 \\ 0 & 2x_3 \end{bmatrix}$$

$\therefore$ the $2x_3$ derivative matrix $\partial y/\partial x$ is as shown above.

2. $x = r\sin\theta\cos\phi \qquad y = r\sin\theta\sin\phi \qquad z = r\cos\theta$

$x = [x; y; z] \Rightarrow [x_1 : x_2 : x_3]$

$y = [r; \theta; \phi] \Rightarrow [y_1 : y_2 : y_3]$

$$\frac{\partial x}{\partial y} = \begin{bmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_2}{\partial x_1} & \frac{\partial x_3}{\partial x_1} \\ \frac{\partial x_1}{\partial x_2} & \frac{\partial x_2}{\partial y_2} & \frac{\partial x_3}{\partial y_3} \\ \frac{\partial x_1}{\partial y_3} & \frac{\partial x_2}{\partial y_3} & \frac{\partial x_3}{\partial y_3} \end{bmatrix}$$

To compute jacobian $\partial x / \partial y$

$\partial x / \partial r = \sin(\theta) * \cos(\phi)$

$\partial x / \partial \theta = r * \cos(\theta) * \cos(\phi)$

$\partial x / \partial \phi = -r * \sin(\theta) * \sin(\phi)$

$\partial y / \partial r = \sin(\theta) * \sin(\phi)$

$\partial y / \partial \theta = r * \cos(\theta) * \sin(\phi)$

$\partial y / \partial \phi = r * \sin(\theta) * \cos(\phi)$

$\partial z / \partial r = \cos \theta$

$\partial z / \partial \theta = -r * \sin \theta$

$\partial z / \partial \phi = 0$

After Assembling this into 3×3 jacobian.

$$\begin{bmatrix} \sin(\theta) * \cos(\phi) & r * \cos(\theta) * \cos(\phi) & -r * \sin(\theta) * \sin(\phi) \\ \sin(\theta) * \sin(\phi) & r * \cos(\theta) * \sin(\phi) & r * \sin(\theta) * \cos(\phi) \\ \cos(\theta) & -r * \sin(\theta) & 0 \end{bmatrix}$$

$\therefore$ The jacobian $\partial x / \partial y$ is the 3×3 matrix

Problem 3:

1. The Hessian of the least square loss is

$L(w) = (\frac{1}{2}) \Sigma (i=1)^n (x\_i^T w - y\_i)^2$

where x_i is input

y_i is the ith target and w is

Parameter vector

we should take 2nd partial derivative

$\partial 2 L / \partial w\_j \partial w\_k = \Sigma (i=1)^n x\_ij x\_ik$

forms. the Hessian matrix is

$H = X^T X$

X is design matrix, each row is a feature vector x_i.

2.

Given,

the first iteration of Newton's method

gives us $w^* = (XX^T)^{-1} XY$

New rule is

$$w \leftarrow w - H^{-1} \nabla L(w)$$

$$= w - (X^T X)^{-1} X^T (Xw - y)$$

as of the definition

$$w \leftarrow (X^T X)^{-1} X^T y$$

$\therefore$ Newton method coverges imediatly in one iteration for linear regression

Hence $w^* = (XX^T)^{-1} Xy$

4. Problem

1. The constrained optimization Problem

$$\min L(W) = \Sigma (f(x_n; W) - t_n)^2$$

Subject to $\|W\|^p \leq Y$

To Convert Lagrangian form

$$L(w, \lambda) = \Sigma (f(x_n; w) - t_n)^2 + \lambda (\|w\|^p - Y)$$

Setting the derivative w.r.t w to 0.

$\therefore$ It gives optimality condition:

$$\nabla L(w, \lambda) = \nabla L(W) + \lambda \nabla \|W\|^p = 0$$

which is same as the optimality condition for

$$\min L(W) = \Sigma (f(x_n; W) - t_n)^2 + \lambda \|W\|^p$$

So the Two Problems are equivalent

At the optimal w, the constraint is satisfied as equally.

$$\|W\|^p = Y$$

Plugging this into the lagrange dual function

$L(w,\lambda) = L(w) + \lambda(Y - Y) = L(w)$

2. $\lambda = Y$ makes the solutions equivalent

for the hyper parameter λ and Y.

1. Y controls the constraint boundary formulation

2. λ controls the regularization strength in the regularized formulation.

3. λ is directly optimized Y is imposed as a constraint.

4. λ and Y play similar roles in controlling model complexity, with λ being the optimized parameter.

5. Problem

1. By gradient update rule and continuity of ∇f

$f(y) \geq f(x) + \nabla f(x)^T (y-x)$

using this with $x = x(k)$, $y = x(k+1)$;

$f(x(k+1)) \geq f(x(k)) + \nabla f(x(k))^T (x(k+1) - x(k))$

GD update $x(k+1) = x(k) - a \nabla f(x(k))$.

$f(x(k+1)) \geq f(x(k)) - a \|\nabla f(f(x(k)))\|^2_2$

$\leq f(x)(k)) - (1 - La/2) a \|\nabla f(x)(k)\|^2_2$.

So f decreases by leas $(a/2) \|\nabla f(x(k))\|^2_2$ Per iteration.

2. By convexity of f:

By Lipschitz condition

$\|\nabla f(x) - \nabla f(y)\| \leq L \|x - y\|$

Applying this to $x = x(k)$ and $y = x(k+1)$,

$f(x(k+1)) \leq f(x(k)) - a \|\nabla f(x)(k))\|^2 +$

$(La^2/2) \|\nabla f(x(k))\|^2$

using $a \leq 1/L$, we get

$$f(x(k+1)) \leq f(x)(k) - (1 - L a/2) a \|\nabla f(x)(k)\|^2$$

$$\leq f(x)(k) - (1/2) a \|\nabla f(x(k))\|^2$$

3. Summing over k iterations and using $f(x) \geq f(x^*)$:

$$\sum_{(k=0)}^{(k-1)} [f(x)(k) - f(x^*)] \leq (1/2a) \|x(0) - x^*\|^2$$

So $f(x(k)) - f(x^*) \leq (1/2ak)) \|x(0) - x^*\|^2$

Therefore, gradient descent ~~conv~~ converges at rate $O(1/k)$. This also proved that convergence rate for gradient descent on convex differentiable functions.