Optimization Pick 5 questions, 1 pts per question

Newton's Method in practice

Due to which reason is Newton's method not often used in machine learning?

Correct Answer

Computing the Hessian is very computationally expensive

Newton's method is simply not effective at finding a solution

Newton's method requires closed-form representations

Newton's method does not work on signals

Least-squares

True or false: Least-squares is a special case of convex optimization.

Correct Answer

True

False

Lagrangian

Suppose we have the following optimization problem

LaTeX: \begin{array}{l} \text{min } f(x) = x^{e^{\frac{1}{3}}}y^{\frac{1}{2}}\\ \text{s.t. } g(x) = 3x^{\frac{1}{3}} + 4y^{\frac{1}{4}}\end{array}min f ( x ) = x e 1 3 y 1 2 s.t. g ( x ) = 3 x 1 3 + 4 y 1 4

Which of the following is the Lagrangian function, LaTeX: \mathcal{L}(x,y,\lambda)L ( x , y , λ )?

Correct Answer

LaTeX: \mathcal{L}(x,y,\lambda) = -4\lambda y^{\frac{1}{4}} + x^{e^{\frac{1}{3}}}y^{\frac{1}{2}} - 3 \lambda

x^{\frac{1}{3}}L ( x , y , λ ) = − 4 λ y 1 4 + x e 1 3 y 1 2 − 3 λ x 1 3

LaTeX: \mathcal{L}(x,y,\lambda) = 4\lambda y^{\frac{1}{4}} - 3\lambda x^{\frac{1}{3}} - x^{e^{\frac{1}{3}}}y^{\frac{1}{2}}L ( x , y , λ ) = 4 λ y 1 4 − 3 λ x 1 3 − x e 1 3 y 1 2

LaTeX: \mathcal{L}(x,y,\lambda) = 4\lambda y^{\frac{1}{4}} + 3x^{\frac{1}{3}} - \lambda x^{e^{\frac{1}{3}}}y^{\frac{1}{2}}L ( x , y , λ ) = 4 λ y 1 4 + 3 x 1 3 − λ x e 1 3 y 1 2

LaTeX: \mathcal{L}(x,y,\lambda) = 4y^{\frac{1}{4}} - 3x^{\frac{1}{3}} - \lambda^2 x^{e^{\frac{1}{3}}}y^{\frac{1}{2}}L ( x , y , λ ) = 4 y 1 4 − 3 x 1 3 − λ 2 x e 1 3 y 1 2

Primal Problem

True or false: The primal problem has a unique solution.

True

Correct Answer

False

Learning rate

Let LaTeX: \eta\_i = f(i)η i = f ( i ) where LaTeX: \eta\_iη i is the learning rate at iteration LaTeX: ii. Which of the following functions is the most suitable to use in gradient descent where LaTeX: \lambdaλ is some constant?

Correct Answer

LaTeX: f(i) = e^{-\lambda i}f ( i ) = e − λ i

LaTeX: f(i) = \lambda if ( i ) = λ i

LaTeX: f(i) = e^{\lambda i}f ( i ) = e λ i

LaTeX: f(i) = -\lambda if ( i ) = − λ i

Newton's Method

Suppose we start Newton's method at a point close to the solution. Will it converge quickly?

Correct Answer

Yes

No, since no one uses Newton's method in practice

No, Newton's method will keep bouncing around the solution

No, Newton's method is slow and expensive to compute

Gradient Descent

Suppose we start gradient descent at a point close to the solution. Should we always assume that we will reach the solution quickly?

Correct Answer

No

Yes, since we are close to the solution

Yes, gradient descent is very effective at finding the solution

Yes, since we move in the opposite direction of the gradient

Solution to a function

The function LaTeX: f\left(x\_1,\:x\_2,\:x\_3\right)\:=\:x^2\_1\:-\:x^2\_2\:+\:x^2\_3\:+\:x\_1x\_2\:+\:x\_1x\_3\:+\:x\_2x\_3\:+\:x\_1\:+\:x\_2\:+\:x\_3\:+\:3f ( x 1 , x 2 , x 3 ) = x 1 2 − x 2 2 + x 3 2 + x 1 x 2 + x 1 x 3 + x 2 x 3 + x 1 + x 2 + x 3 + 3 has (do not consider values at LaTeX: \pm\infty± ∞)

Correct Answer

A saddle point

A minimum

A maximum

None of the above

Hessian

The Hessian of the following function:

LaTeX: f\left(x\_1,\:x\_2,\:x\_3\right)\:=\:x^2\_1x\_2\:+\:x^2\_2x\_3\:+\:x\_3^3\:+\:2x\_1x\_3\:+\:x\_2\:+\:6f ( x 1 , x 2 , x 3 ) = x 1 2 x 2 + x 2 2 x 3 + x 3 3 + 2 x 1 x 3 + x 2 + 6 at (1, 1, 1) is

Correct Answer

Positive semidefinite

Negative semidefinite

Indefinite

Positive definite

Negative definite

Derivatives

Which of the following is/are true?

Correct Answer

The first and second derivative of the objective function are used to determine if we have reached an optimal point

Correct Answer

The Hessian is the matrix of second-order partial derivatives of a function

Correct Answer

The gradient of f(x, y) is the a vector pointing in the direction of the steepest slope at that point

Direct optimization methods are useful when the objective function is not differentiable

Deterministic Representation Pick 5 questions, 1 pts per question

Best Bases

According to what we discussed in the class, which of the following alternatives is the best choice for a bases in LaTeX: \mathbb{R}^3R 3.

LaTeX: \vec{b\_{1}} = \begin{bmatrix}

3\\

1

\\

0

\end{bmatrix} ,

\vec{b\_{2}} = \begin{bmatrix}

1\\

5

\\

7

\end{bmatrix},

\vec{b\_{3}} = \begin{bmatrix}

9\\

2

\\

7

\end{bmatrix}

b 1 → = [ 3 1 0 ] , b 2 → = [ 1 5 7 ] , b 3 → = [ 9 2 7 ]

LaTeX: \vec{b\_{1}} = \begin{bmatrix}

9\\

0

\\

0

\end{bmatrix} ,

\vec{b\_{2}} = \begin{bmatrix}

0\\

1

\\

0

\end{bmatrix},

\vec{b\_{3}} = \begin{bmatrix}

0\\

0

\\

5

\end{bmatrix},

\vec{b\_{4}} = \begin{bmatrix}

1\\

0

\\

5

\end{bmatrix}

b 1 → = [ 9 0 0 ] , b 2 → = [ 0 1 0 ] , b 3 → = [ 0 0 5 ] , b 4 → = [ 1 0 5 ]

LaTeX: \vec{b\_{1}} = \begin{bmatrix}

3/7\\

-6/7

\\

2/7

\end{bmatrix} ,

\vec{b\_{2}} = \begin{bmatrix}

2/7\\

3/7

\\

6/7

\end{bmatrix},

\vec{b\_{3}} = \begin{bmatrix}

6/7\\

2/7

\\

-3/7

\end{bmatrix}b 1 → = [ 3 / 7 − 6 / 7 2 / 7 ] , b 2 → = [ 2 / 7 3 / 7 6 / 7 ] , b 3 → = [ 6 / 7 2 / 7 − 3 / 7 ]

LaTeX: \vec{b\_{1}} = \begin{bmatrix}

3\\

2

\\

6

\end{bmatrix} ,

\vec{b\_{2}} = \begin{bmatrix}

1\\

11

\\

7

\end{bmatrix},

\vec{b\_{3}} = \begin{bmatrix}

16\\

52

\\

52

\end{bmatrix}b 1 → = [ 3 2 6 ] , b 2 → = [ 1 11 7 ] , b 3 → = [ 16 52 52 ]

Correct Answer

3

1

2

4

Complex Exponentials

Which of the following is true about complex exponential bases and DFT?

Correct Answer

Complex exponential bases take phase into consideration

Complex exponential bases are not orthogonal

Phases we trying to optimize are included in basis rather than weight

DFT is an algorithm which performs FFT really fast

Windows

Why is a sound signal processed in small windows?

Correct Answer

The properties of a signal are considered stationary

It is computationally more effective to use small windows

Small windows allow complex exponentials to become real

Small windows give us more frequencies

Properties of Complex Exponentials

Which of the following is/are true?

Correct Answer

The complex exponentials with frequencies equally spaced from L/2 are complex conjugates

Correct Answer

Complex exponentials are well-behaved

Correct Answer

Complex exponentials allow the phase to be estimated

The inverse of a complex matrix is always equal to its Hermitian

Checkerboards as bases

Why are checkerboards used as bases?

Correct Answer

Checkerboards capture changes with varying speeds

Correct Answer

Checkerboards are orthogonal

Checkerboards act as band-pass filters to avoid aliasing

Checkerboards allow us to capture round edges

Checkerboards & Sinusoids

Which of the following is false about checkerboards and sinusoid bases?

Correct Answer

Checkerboards and sinusoid bases are good because they can both represent sharp corners and round curves

Checkerboard bases are orthogonal to each other

Sinusoidal bases are orthogonal to each other

To represent any signal using sinusoid bases we need to consider a phase term in the basis

Bases

Let LaTeX: VV be a LaTeX: nn dimensional space, and LaTeX: \mathcal{B} = \{{\bf v}\_1,...,{\bf v}\_n\}B = { v 1 , . . . , v n } a basis of LaTeX: VV. Then, for any vector LaTeX: {\bf x}x in LaTeX: VV, there exists LaTeX: \lambda\_1,...,\lambda\_nλ 1 , . . . , λ n, such that LaTeX: {\bf x} = \lambda\_1 {\bf v}\_1 +...+ \lambda\_n {\bf v}\_nx = λ 1 v 1 + . . . + λ n v n.

Hence, we can define the length of LaTeX: {\bf x}x with respect to LaTeX: \mathcal{B}B as follows LaTeX: \| {\bf x}\|\_{\mathcal{B}} = \sqrt{\sum\_{i=1}^n \lambda\_i^2}‖ x ‖ B = ∑ i = 1 n λ i 2

Under which conditions this expression is equal to LaTeX: \sqrt{\sum\_{i=1}^n x\_i^2}∑ i = 1 n x i 2:

This is always true

LaTeX: \mathcal{B}B must be an orthonormal basis

This is never true

LaTeX: \mathcal{B}B must be an orthogonal basis

Correct Answer

LaTeX: \mathcal{B}B must be an orthonormal basis

This is always true

This is never true

LaTeX: \mathcal{B}B must be an orthogonal basis

Deterministic Representations

Why do we learn deterministic representations from data?

Correct Answer

Allows data to be efficiently stored

Correct Answer

Allows data compression

Correct Answer

Enables downstream machine learning applications

Enables Fourier algorithms like FFT to run faster

Fourier Transform

Suppose you run a DFT function from your favorite Python library, but the function only returns LaTeX: \frac{L}{2}L 2 Fourier transform weights given a real input signal of length LaTeX: LL samples. This is unfortunate because the quiz question you are trying to answer asks you to compute LaTeX: S\left(\frac{L}{2}+3\right)S ( L 2 + 3 ), the LaTeX: \left(\frac{L}{2}+3\right)^{th}( L 2 + 3 ) t h Fourier transform weight. Given this information, how should you proceed to answer the quiz question?

Correct Answer

LaTeX: S\left(\frac{L}{2}+3\right)=S^{\ast}\left(\frac{L}{2}-3\right)S ( L 2 + 3 ) = S ∗ ( L 2 − 3 )

You cannot answer this question without more information; find a new Python library

LaTeX: S\left(\frac{L}{2}+3\right)=-S\left(\frac{L}{2}-3\right)S ( L 2 + 3 ) = − S ( L 2 − 3 )

LaTeX: S\left(\frac{L}{2}+3\right)=S\left(\frac{L}{2}-3\right)S ( L 2 + 3 ) = S ( L 2 − 3 )

STFT

Notice in homework 1 that when the "n\_fft" parameter of the "stft" function is set to return a 2048-point FFT, the output matrix only contains 1025 frequency channels. Why is this the case?

Correct Answer

The magnitude of channels 1026-2048 is equivalent to that of channels 2-1024

The phase of channels 1026-2048 is equivalent to that of channels 1-1023

A 2048-point FFT only returns 1025 weights by definition; the other weights don't exist

The magnitude of channels 1026-2048 is equivalent to that of channels 1-1023