Announcements

- Final Exam
  - Thursday October 14 from 1-4pm in this room.

- No class Wednesday!

- Quiz 3:
  - Last 15 minutes of this class.

Today: Exam Review

- Linear Algebra
- Calculus
- Optimization
- Probability and Statistics

Linear Algebra: Inner Product vs Outer Products (Lecture 3)

Let $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ be column vectors.

- Inner Product

$$\mathbf{u}^T\mathbf{v} = \mathbf{v}^T\mathbf{u} = \sum_{i=1}^{n} u_i v_i$$

$$\mathbf{u}^T\mathbf{u} = ||\mathbf{u}||_2^2$$

- Outer Product

$$\mathbf{u}\mathbf{v}^T = \underbrace{\begin{bmatrix} | & | & & | \\ v_1\mathbf{u} & v_2\mathbf{u} & \ldots & v_n\mathbf{u} \\ | & | & & | \end{bmatrix}}_{\text{nxn}}$$

Linear Algebra: Span and Rank (Lecture 3)

- **Span**

  Set of all linear combination of a set of vectors

  Given a set of vectors $\mathcal{S} = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3\}$, where $\boldsymbol{v}_i \in \mathbb{R}^M \ \forall\, i$

  $$span(\mathcal{S}) = \{w_1\boldsymbol{v}_1 + w_2\boldsymbol{v}_2 + w_3\boldsymbol{v}_3 \mid w_i \in \mathbb{R}\}$$

- **Rank**
  - The rank of a matrix is the dimension of the span of its columns.
  - Or equivalently, the number of linearly independent columns, eg...

Vector $\mathbf{v}_3$ is linearly independent from vectors $\mathbf{v}_1$ and $\mathbf{v}_2$ if there do not exists weights $w_1, w_2 \in \mathbb{R}$ such that $\mathbf{v}_3 = w_1\mathbf{v}_1 + w_2\mathbf{v}_2$

Span and Rank Practice

- What are the ranks of these matrices?

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$A = \begin{bmatrix} 1 & 2 & 1 \\ 1 & 2 & 2 \\ 1 & 2 & 1 \end{bmatrix}$$

Span and Rank Practice

- What is the rank of this matrix?

$$A = \begin{bmatrix} 1 & 0 & 1 & 2 \\ 1 & 1 & 0 & 2 \\ 0 & 1 & -1 & 0 \end{bmatrix}$$

a) 1

b) 2

c) 3

d) 4

Span and Rank Practice

Let $\mathbf{u} \in \mathbb{R}^n$ and $\mathbf{v} \in \mathbb{R}^m$ where $n < m$. What is the rank of $\mathbf{u}\mathbf{v}^T$?

a) 1

b) n

c) m

d) uv^T is a scalar.

e) Cannot be determined without knowing exact values of **u** and **v.**

Linear Algebra: Things to Know! (Not Exhaustive)

- Inner products vs outer products
- What is the shape that results after doing different matrix operations.
- Span and rank
- Transpose Rules
- Matrix distributive property rules
- Linear systems, how many solutions will there be in different cases

Calculus: Matrix Derivatives (Lecture 6)

- In every scenario, we are just taking a bunch of partial derivatives.

- How do we determine how many scalar derivatives there will be?

For the derivative $\frac{\partial f}{\partial x}$, the total number of derivatives is...

$$\text{Total Number of Derivatives} = (\text{Number of entries in } x) \times (\text{Number of entries in the output of } f)$$

- Basically we see how each entry in the output changes with respect to each entry in x.

Matrix Derivatives Practice

- How many partial derivatives are there for…

1) $\frac{\partial f}{\partial A}$ where $f(\mathbf{x}, A) = A\mathbf{x}$ and $A \in \mathbb{R}^{a \times b}$ and $\mathbf{x} \in \mathbb{R}^{b}$.

2) $\frac{\partial f}{\partial A}$ where $f(A, B, C) = BAC$ and $A \in \mathbb{R}^{m \times n}$, $B \in \mathbb{R}^{b \times m}$, and $C \in \mathbb{R}^{n \times c}$.

Matrix Derivatives Practice

- How many partial derivatives are there for…

$$\frac{\partial f}{\partial \mathbf{x}} \text{ where } f(\mathbf{x}) = \mathbf{x} \text{ and } \mathbf{x} \in \mathbb{R}^b.$$

a) 1

b) b

c) b^2

d) b^3

Calculus: Numerator format (Lecture 6)

- Numerator format is the particular way we order the partial derivatives.

$$\text{Let } \mathbf{y} = f(\mathbf{x}) \text{ where } \mathbf{y} \in \mathbb{R}^m \text{ and } \mathbf{x} \in \mathbb{R}^n.$$

$$\frac{\partial \mathbf{y}}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial \mathbf{y}_1}{\partial \mathbf{x}_1} & \cdots & \frac{\partial \mathbf{y}_1}{\partial \mathbf{x}_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial \mathbf{y}_m}{\partial \mathbf{x}_1} & \cdots & \frac{\partial \mathbf{y}_m}{\partial \mathbf{x}_n} \end{bmatrix}$$

- When the output is a scalar this is a row vector.
- When the input is a scalar this is a column vector.
- When both are vectors this is the *Jacobian*

Let $\mathbf{y} = f(\mathbf{x})$ where $\mathbf{y} \in \mathbb{R}$ and $\mathbf{x} \in \mathbb{R}^n$. Assuming numerator format, how does $\frac{\partial \mathbf{y}}{\partial \mathbf{x}}$ and $\nabla_{\mathbf{x}} f(\mathbf{x})$ relate to each other?

a) $\nabla_{\mathbf{x}} f(\mathbf{x}) = \frac{\partial \mathbf{y}}{\partial \mathbf{x}}$

b) $\nabla_{\mathbf{x}} f(\mathbf{x}) = \left(\frac{\partial \mathbf{y}}{\partial \mathbf{x}}\right)^T$

c) They have no relation.

d) They are only sometimes equal.

Calculus: Things to Know! (Not Necessarily Exhaustive)

- Numerator Form
- Knowing number of entries/shape of derivative.
- Standard matrix derivative rules.
- Chain Rule and Multivariate Chain Rule (Lecture 6 Slides 23-onward)

Optimization: Linear Functions (Lecture 4)

- Properties of linear function:

  Linear function

  If $f(\boldsymbol{x})$ is linear, then for any $\boldsymbol{u}, \boldsymbol{v}$:

  - $f(\boldsymbol{u} + \boldsymbol{z}) = f(\boldsymbol{u}) + f(\boldsymbol{z})$
  - $f(\alpha \boldsymbol{z}) = \alpha f(\boldsymbol{z}) \quad \forall \alpha$
  - $f(\alpha \boldsymbol{u} + (1 - \alpha)\boldsymbol{z}) = \alpha f(\boldsymbol{u}) + (1 - \alpha) f(\boldsymbol{z}) \quad \forall \alpha$

- Is this function linear?
  - f(x) = |x|

Is this function linear?

$$f(x) = c \text{ where } c \in \mathbb{R}$$

a) Yes

b) No

c) It depends

Optimization: Convex Functions (Lecture 4)

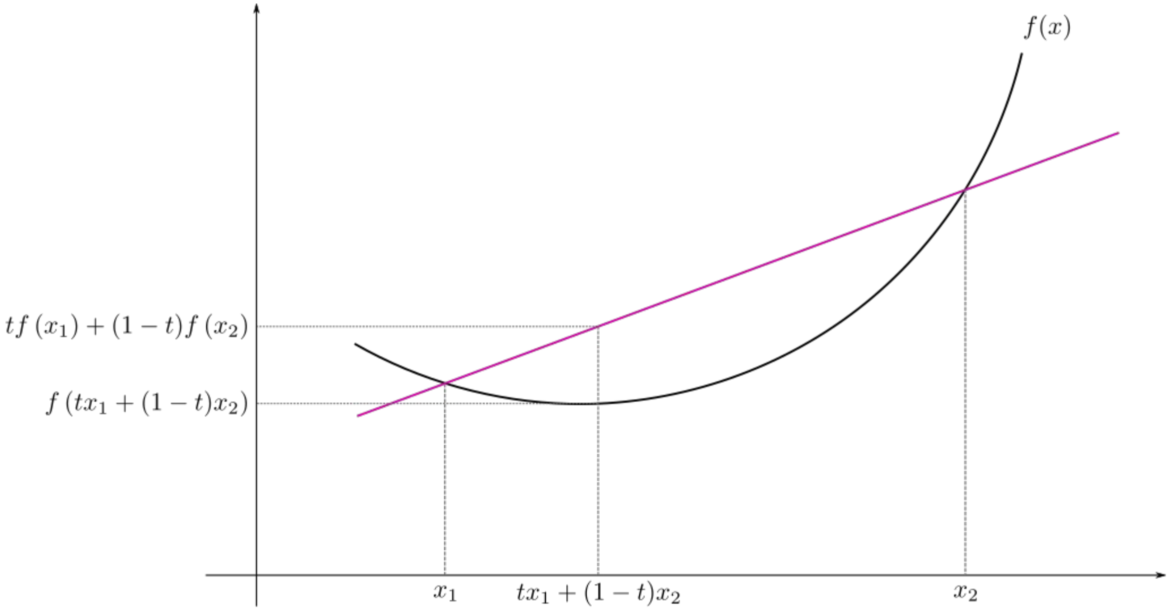
Convex function

If $f(\boldsymbol{x})$ is convex, then:

- $f(\alpha\boldsymbol{x} + (1-\alpha)\boldsymbol{z}) \leq \alpha f(\boldsymbol{x}) + (1-\alpha) f(\boldsymbol{z}) \quad \forall\, 0 \leq \alpha \leq 1$

- Critical points for convex functions are global minima.

- Which objective functions that we have looked at are convex?



Optimization: Things to Know! (Not Necessarily Exhaustive)

- Linear Functions
- Convex Functions
- When are critical points global minima
- How to form Lagrangian. Why is the Lagrangian useful?
- What the MSE objective is in both the scalar and matrix form.

Probability: Chain Rule and Marginalization

- The Chain Rule

$$P(A,B) = P(A|B)P(B)$$

  - More interesting example:

$$P(X_1, X_2, X_3, X_4, X_5 | Y_1, Y_2, Y_3) = P(X_1, X_3, X_5 | X_2, X_4, Y_1, Y_2, Y_3) P(X_2, X_4 | Y_1, Y_2, Y_3)$$

- Marginalization

$$P(X) = \sum_y P(X, Y = y)$$

Probability: A good strategy for some of these problems is to first get to the joint.

Suppose we want to compute $P(A|B)$ but only have the probability tables for $P(B|A,C)$ and $P(A,C)$...

$$P(A|B) = \frac{P(A,B)}{P(B)}$$

$$= \frac{\sum_c P(A,B,C=c)}{\sum_a \sum_c P(A=a,B,C=c)}$$

$$= \frac{\sum_c P(B|A,C=c)P(A,C=c)}{\sum_a \sum_c P(B|A=a,C=c)P(A=a,C=c)}$$

Statistics vs Probability

- Probability assumes we know how data is generated and tries to predict likelihood of future events happening.

- Statistics takes data that has already happened and tries to infer information about how the data was generated.

Recipe for Estimation

(Lecture 10, slide 18)

MLE

1. Formulate the likelihood, $p(\mathcal{D} \mid \theta)$
2. Set objective $J(\theta)$ equal to negative log likelihood

$$\mathrm{J}(\theta) = -\log p(\mathcal{D} \mid \theta)$$

3. Compute derivative of objective, $\partial J / \partial \theta$
4. Find $\hat{\theta}$, either
   a. Set derivate equal to zero and solve for θ
   b. Use (stochastic) gradient descent to step towards better θ

Probability and Statistics: Things to Know! (Not Necessarily Exhaustive)

- Chain Rule for Probability
- Bayes Rule
- Marginalization
- Independence
- Conditional Independence
- Expected Value
- Variance
- MLE
- Log properties for forming negative log likelihood/taking derivatives