Outline

• Linear algebra – geometric interpretation
• Probability – inequalities and bounds
• Some interesting stuff
Linear Algebra

Vectors, vector spaces, matrices, SVD
Vectors

• \( \mathbf{x} = (x_1, x_2, ..., x_d) \in \mathbb{R}^d \) (each \( x_i \) is a component)
  • A point in \( d \)-dimensional space

• Norm or magnitude \( \| \mathbf{x} \| = (\mathbf{x}^T \mathbf{x})^{1/2} = (x_1^2 + x_2^2 + ... + x_d^2)^{1/2} \)
  • Length of the vector (Pythagorean theorem)

• Zero vector (norm zero), unit vector (norm one)

• Inner product \( \langle \mathbf{x}, \mathbf{y} \rangle = x_1 y_1 + ... x_d y_d \)
  • Result is a scalar
  • \( \| \mathbf{x} \| = (\langle \mathbf{x}, \mathbf{x} \rangle)^{1/2} \)
  • \( \langle \mathbf{x}, \mathbf{y} \rangle = 0 \) implies \( \mathbf{x} \perp \mathbf{y} \)
Vector spaces

• Space where vectors live

• Formally, a collection of vectors which is closed under linear combination
  • If \{x, y\} are in the space, so is ax+by for any scalars a, b ∈ R
  • Should always contain zero vector

• Examples: \{0\}, \mathbb{R}^d, the line x = 3y in \mathbb{R}^2
Span and basis

• A set of vectors is said to span a vector space if one can write any vector in the vector space as a linear combination of the set
  
  \[ \{x_1, x_2, ..., x_n\} \text{ span the space } \{ \sum a_i x_i \mid a_i \in \mathbb{R} \} \]

• This set is called the basis set

• Examples
  
  • The vectors \{ (0,1), (1,0) \} span \( \mathbb{R}^2 \)
  
  • \{ (1, 1) \} spans \( x=y \) which is a subspace of \( \mathbb{R}^2 \)
  
  • The vector \{ (0,1), (0,1), (1,1) \} also span \( \mathbb{R}^2 \)
Linear independence and orthonormality

• Linear independence – a notion to remove redundancy in the basis
  • \{x_1, x_2, ..., x_n\} are linearly independent iff the only solution to \sum a_ix_i = 0 is \(a_1 = a_2 = ... = a_n = 0\).
  • Cannot express any vector \(x_i\) as a linear combination of the others

• Dimensionality of a vector space is the maximum number of linearly independent basis vectors

• Orthonormal basis
  • \{x_1, x_2, ..., x_n\} is orthonormal basis if \(\langle x_i, x_j \rangle = 1\) if \(i=j\) and 0 otherwise
  • Coordinate axes for the vector space

• Example: The basis \{(0, 1), (1,1)\} for \(R^2\) is linear independent but not orthonormal.
Matrices

- Operator which transforms vectors from one vector space to another
  - $y = Ax$
- The operator is linear, that is
  \[ A(ax + by) = a(Ax) + b(Ay) \]
- The result of applying the operator is a linear combination of the column vectors
  - Thus, $Ax = b$ has an exact solution iff $b$ is in the column space of $A$
- Eigen vectors of $A$ are the special vectors are the special vectors $x$ which satisfy
  \[ Ax = \lambda x \text{ for some } \lambda \]
  - $\lambda$ is called the eigen value and $x$ is the eigen vector
- How do we visualize the transformation geometrically?
Visualizing the matrix operator – special cases

- **Identity matrix**
  - Square matrix with diagonal elements 1 and non-diagonal elements 0
  - The transformed vector $Ax$ is same $x$

- **Diagonal matrix**
  - Square matrix with non-diagonal elements 0
  - $i^{th}$ component in $Ax$ is a scaled version of $x_i$ (scaling = $A_{ii}$)

- **Orthonormal (or rotation) matrix**
  - Matrix whose columns $\{a_1, a_2, ..., a_n\}$ are such that $<a_i, a_j> = 1$ if $i=j$ and 0 otherwise. That is, $A^T A = I$
  - Rotates the vector
  - Preserves norms $\|Ax\| = \|x\|$ (why?)
General case – Singular Value Decomposition

• We have a rectangular matrix $A \in \mathbb{R}^{m \times n}$
• It can be decomposed as
  \[ A = U D V^T \]
  
  • $U$ and $V$ are orthonormal, i.e., $U^T U = V^T V = I$ and $D$ is a diagonal matrix containing singular values
    • Number of non-zero diagonal elements in $D = \text{rank of } A$
• Provides a nice way to understand the operator $A$
  • Rotation in $n$-dimensional space, scaling, rotation in $m$-dimensional space
• Can be computed in $O(\min\{mn^2, m^2n\})$ time (or better using fast matrix multiplication)
Computation of SVD

• Let $m>n$, i.e., $A$ is a skinny matrix. How to compute SVD of $A$ in $O(mn^2)$ time?

• Step 1: Compute $A^T A$ in $O(mn^2)$ time.

• Step 2: Get eigenvalue decomposition of $A^T A$ in $O(n^3)$ or better. Why do this?
  • If the SVD of $A$ is $UDV^T$, then $A^T A = VDU^T UD V^T$
  • That is, the eigenvalues of $AA^T$ are the square of the singular values of $A$ and the eigenvectors are the right singular space

• Step 3: $U = AVD^{-1}$ in $O(mn^2)$ time.
Example problem 1

• If singular values of $A \in \mathbb{R}^{n \times n}$ all lie in $[a, b]$, prove that
  \[ a\|x\| \leq \|Ax\| \leq b\|x\| \]

Solution:
• Let $A = UDV^T$
• $\|Ax\| = \|UDV^T x\|$
• Let $y = V^T x$. (note: $\|y\| = \|x\|$)
  • We can do this because we prove this for every $x$
• $\|Ax\| = \|UDy\| = \|Dy\|$
• As singular values lie in $[a, b]$, $a\|y\| \leq \|Dy\| \leq b\|y\|$
Example problem 2

• Prove that Frobenius norm of a matrix \( \|A\|_F = (\sum_i \sum_j A_{ij}^2)^{1/2} \) is always greater than or equal to the operator norm \( \|A\|_2 = \sup_x \|Ax\|/\|x\| \). Solution:

Solution:

• Let \( x = \sum_j c_j e_j \) for coefficients \( c_1, .. c_d \)
• Let \( \|x\|_2 = 1 \). Then, \( \sum_j |c_j|^2 = 1 \)
• \( \|Ax\|_2^2 = \| \sum_j c_j A e_j \|_2^2 \)
• By triangle inequality, this is \( \leq \left( \sum_j |c_j| \|A e_j\|_2 \right)^2 \)
• Which is \( \leq \left( \sum_j |c_j|^2 \right) \left( \|A e_j\|_2^2 \right) \) by Cauchy-Schwarz inequality
• Which is \( \|A e_j\|_2^2 = \|A\|_F \)
Probability

Useful inequalities
Expectation and variance

• Let $X$ be a random variable
• Expectation $E[X] = \sum_j P(X=j) \cdot j$ (discrete)
• Variance $\text{Var}[X] = E[(X-E[X])^2] = E[X^2] - E[X]^2$
• In general, $k^{th}$ order moment is $E[|X-E[X]|^k]$
Markov inequality

• For a non-negative random variable $X$ and non-negative $t$,
  \[ \Pr[X \geq t] \leq \frac{E[X]}{t} \]

Proof:
• We’ll show for continuous r.v, but proof is similar for discrete r.v
• $E[X] = \int_0^\infty x \, p(x) \, dx = \int_0^t x \, p(x) \, dx + \int_t^\infty x \, p(x) \, dx$
• $E[X] \leq \int_t^\infty x \, p(x) \, dx$
• $\leq t \cdot \int_t^\infty p(x) \, dx = t \cdot \Pr[X \geq t]$
Chebyshev inequality

• Let $\mu = \mathbb{E}[X]$ and $\sigma^2 = \text{Var}[X]$. Then,
  \[ \Pr[|X-\mu| \geq t] \leq \frac{\sigma^2}{t^2} \]

**Proof:**

• $\Pr[|X-\mu| \geq t] = \Pr[|X-\mu|^2 \geq t^2]$

• By Markov inequality, $\Pr[|X-\mu|^2 \geq t^2] \leq \mathbb{E}[|X-\mu|^2]/t^2 = \sigma^2/t^2$
Chernoff bound

• For independent random variables $X_1, X_2, \ldots X_n$, with $X = \sum_i X_i$
  \[
  \Pr[X \geq a] \leq \min_{t \geq 0} e^{-ta} \prod_i E[e^{tX_i}]
  \]
  \[
  \Pr[X \leq a] \leq \min_{t \geq 0} e^{ta} \prod_i E[e^{-tX_i}]
  \]

Proof:

• Key idea: Apply Markov inequality on $e^{tx}$
  \[
  \Pr[X \geq a] = \Pr[e^{tx} \geq e^{ta}] \leq e^{-ta} E[e^{tx}]
  \]

• By independence, this is $e^{-ta} \prod_i E[e^{tX_i}]$

• This is true for every positive $t$, so take infimum to get the best bound
Chernoff bound (i.i.d Bernoulli)

• For independent Bernoulli random variables $X_1, X_2, .. X_n$ each having probability $p$ of being equal to 1, if $X$ is the sum $\sum_i X_i$,

$$\Pr(X > (1 + \delta)\mu) < \left(\frac{e^{\delta}}{(1 + \delta)^{1+\delta}}\right)^\mu$$

• A more useful but loose bound is:

$$\Pr(X \geq (1 + \delta)\mu) \leq e^{-\frac{\delta^2 \mu}{3}}, \quad 0 \leq \delta \leq 1$$
Interesting stuff

Just one this time...
Hadamard matrix–vector product in $O(n \log n)$

- Let $H_k$ be the Hadamard matrix with $2^k$ rows and columns
- Observe that $H_k = \begin{bmatrix} H_{k-1} & H_{k-1} \\ H_{k-1} & -H_{k-1} \end{bmatrix}$
- Let $x$ be $(x_u, x_l)$ – the upper and lower parts contain $n/2$ entries each
- Then, $H_k x = \begin{bmatrix} H_{k-1} x_u + H_{k-1} x_l \\ H_{k-1} x_u - H_{k-1} x_l \end{bmatrix}$
- Once $H_{k-1} x_l$ and $H_{k-1} x_u$ have been computed in $T(n/2)$ time, we perform $O(n)$ element wise addition/subtraction to solve the original problem
- Thus, $T(n) = 2T(n/2) + O(n)$ which gives $O(n \log n)$ time complexity