CME / EE 103 Review Guide

Since the textbook and slides are very comprehensive, this review guide is designed to highlight some of the most important concepts of the course in a concise manner. Each chapter is summarized in less than a page, and many non-essential concepts were omitted. We hope this will help you prepare for the quiz exam.

1 Vectors

- We can write a vector in one of three ways

\[ a = (1, 7, 3) = \begin{bmatrix} 1 \\ 7 \\ 3 \end{bmatrix} = \begin{bmatrix} 1 \\ 7 \\ 3 \end{bmatrix} \]

- An \( n \)-vector is a vector with \( n \) entries (sometimes, we say this vector has length \( n \), but in some contexts length can mean other things). The \( i \)th entry of the \( n \)-vector \( x \) is denoted by \( x_i \).

  **Caution:** Sometimes \( x_i \) refers to the \( i \)th vector, instead of the \( i \)th entry of the vector \( x \). You will always be able to determine this from the context of the question.

- If \( a \) is a \( k \)-vector and \( b \) is a \( p \)-vector, we can define a block vector \( x = (a, b) \) of length \( k + p \).

- If \( x \) is an \( n \)-vector, then \( x_{i:j} = (x_i, ..., x_j) \) is a vector of length \( j - i + 1 \).

- A vector is called **sparse** if it has very few non-zero entries. The function \( \text{nnz}(x) \) returns the number of non-zero entries in \( x \). If \( \text{nnz}(x) \) is much smaller than the number of entries in \( x \), then \( x \) is sparse.

- Both 0 and 1 are vectors of undefined length, and we can generally infer their length from the context of the question.

- A **unit vector** \( e_i \) is a vector with one in its \( i \)th entry and zero in every other entry.

- The **inner product** \( a^T b \) between two \( n \)-vectors \( a \) and \( b \) is

\[ a^T b = a_1 b_1 + a_2 b_2 + \cdots + a_n b_n \]

- Some important properties of the inner product include

  **Commutativity:** \( a^T b = b^T a \)

  **Linearity:** \( a^T (\alpha \cdot b + \beta \cdot c) = \alpha (a^T b) + \beta (a^T c) \) (where \( \alpha, \beta \) are scalars and \( a, b, c \) are \( n \)-vectors)

- A **linear combination** of \( n \)-vectors \( x_1, ..., x_k \) is a vector \( y \) defined by

\[ y = \alpha_1 x_1 + \cdots + \alpha_k x_k \]

where \( \alpha_1, ..., \alpha_k \) are scalars.
2 Linear functions

- When we define a function \( f : \mathbb{R}^n \to \mathbb{R} \), we mean that the function \( f(x) \) has \( n \)-vectors as inputs, and scalars as outputs.
- A function \( f : \mathbb{R}^n \to \mathbb{R} \) is a linear function if and only if it satisfies
  \[
  f(\alpha x + \beta y) = \alpha f(x) + \beta f(y)
  \]
  for all \( n \)-vectors \( x, y \) and all scalars \( \alpha, \beta \).
- Equivalently, \( f \) is a linear function if and only if it \( f(x) = c^T x \) for some \( n \)-vector \( c \). The entries of \( c \) cannot depend on \( x \).
- If we know \( f \) is linear, the entries of \( c \) must be
  \[
  c = \begin{bmatrix}
  f(e_1) \\
  \vdots \\
  f(e_n)
  \end{bmatrix}
  \]
  be sure to understand why.

- A function \( f : \mathbb{R}^n \to \mathbb{R} \) is an affine function if and only if it satisfies
  \[
  f(\alpha x + \beta y) = \alpha f(x) + \beta f(y)
  \]
  for all \( n \)-vectors \( x, y \) and all scalars \( \alpha, \beta \) such that \( \alpha + \beta = 1 \).
- All linear functions are affine, but not all affine functions are linear.
- Equivalently, \( f \) is an affine function if and only if it \( f(x) = c^T x + b \) for some \( n \)-vector \( c \) and some scalar \( b \). The entries of \( c, b \) cannot depend on \( x \).
- If we know \( f \) is affine, we know \( b = f(0) \) and the entries of \( c \) must be
  \[
  c = \begin{bmatrix}
  f(e_1) - f(0) \\
  \vdots \\
  f(e_n) - f(0)
  \end{bmatrix}
  \]
  be sure to understand why.
3 Norm and distance

- The norm of an $n$-vector $x$ is

$$\|x\| = \sqrt{x^T x} = \sqrt{x_1^2 + \ldots + x_n^2}$$

- Some important properties of the norm include

  - **Absolute homogeneity**: $\|\alpha x\| = |\alpha|\|x\|$ where $\alpha$ is a scalar.
  
  - **Cauchy-Schwarz**: $|x^T y| \leq \|x\|\|y\|$ (see the book for a proof)
  
  - **Triangle inequality**: $\|x + y\| \leq \|x\| + \|y\|$  

- The distance between two $n$-vectors $x$ and $y$ is $\|x - y\|$.

- Some definitions you should know and understand

  - **Root mean square**:

$$\text{rms}(x) = \|x\|/\sqrt{n}$$

  where $n$ is the number of entries in $x$.

  - **Average**:

$$\text{avg}(x) = \frac{1^T x}{n}$$

  where $n$ is the number of entries in $x$.

  - **Angle**:

$$\theta = \cos^{-1} \left( \frac{x^T y}{\|x\|\|y\|} \right)$$

- For the following definitions, let $\tilde{x} = x - \text{avg}(x)1$. The vector $\tilde{x}$ is often called the *demeaned* version of $x$.

  - **Standard deviation**:

$$\text{std}(x) = \text{rms}(\tilde{x})$$

  - **Correlation**:

$$\text{corr}(x, y) = \frac{\tilde{x}^T \tilde{y}}{\|\tilde{x}\|\|\tilde{y}\|}$$

Using Cauchy-Schwarz, we know $-1 \leq \text{corr}(x, y) \leq 1$. 

3
4 Clustering

- The best resource for understanding the k-means algorithm is the k-means visualizer located here: [http://stanford.edu/class/ee103/visualizations/kmeans/kmeans.html](http://stanford.edu/class/ee103/visualizations/kmeans/kmeans.html)
- Some rough psuedocode for the k-means algorithm:
  1. Choose $k$ initial centroids $z_1, \ldots, z_k$.
     
     Repeat until (2) and (3) until convergence
  2. For each datapoint $x_j$, compute distances $\|x_j - z_1\|, \ldots, \|x_j - z_k\|$. Find the closest centroid, i.e. choose $i$ to minimize $\|x_j - z_i\|$. Assign $x_j$ to the group $G_i$.
  3. For each group $G_i$, let $x_{j_1}, \ldots, x_{j_{n_i}}$ be the vectors assign to group $G_i$. Set $z_i$ to be the mean of these vectors, i.e.

$$z_i = \frac{x_{j_1} + \cdots + x_{j_{n_i}}}{n_i}$$
5 Linear independence

• Recall a linear combination of n-vectors \(x_1, \ldots, x_k\) is a vector \(y\) defined by
  
  \[ y = \alpha_1 x_1 + \cdots + \alpha_k x_k \]

  where \(\alpha_1, \ldots, \alpha_k\) are scalars.

• A set of vectors \(a_1, \ldots, a_k\) are called linearly dependent if
  
  \[ \beta_1 a_1 + \cdots + \beta_k a_k = 0 \]

  for some \(\beta_1, \ldots, \beta_k\) that are not all zero.

  – Another way of phrasing this is saying a set of vectors are linearly dependent if the zero vector can be formed as a linear combination of these vectors.

  – Caution: Linear dependence is a property of a set or collection of vectors, not of an individual vector.

• A set of vectors \(a_1, \ldots, a_k\) are called linearly independent if they are not linearly dependent. Equivalently, the expression
  
  \[ \beta_1 a_1 + \cdots + \beta_k a_k = 0 \]

  only holds when \(\beta_1 = \cdots = \beta_k = 0\).

  – If \(x = \beta_1 a_1 + \cdots + \beta_k a_k\) and \(a_1, \ldots, a_k\) are linearly independent, then the coefficients \(\beta_1, \ldots, \beta_k\) used to form \(x\) are unique.

• A basis is a set of \(n\) linearly independent \(n\)-vectors.

  – If \(a_1, \ldots, a_n\) are a basis, then any \(n\)-vector can be written as a linear combination of \(a_1, \ldots, a_n\).

• The independence-dimension inequality states that any collection of \(n + 1\) or more \(n\)-vectors is linearly dependent.

• A collection of vectors \(a_1, \ldots, a_k\) is orthogonal if \(a_i^T a_j = 0\) whenever \(i \neq j\).

  – Orthogonal vectors are always linearly independent.

• A collection of vectors \(a_1, \ldots, a_k\) is orthonormal if it is orthogonal and \(||a_i|| = 1\) for \(i = 1, \ldots, k\).

  – If \(x\) is a linear combination of orthonormal vectors \(a_1, \ldots, a_k\) defined by \(x = \beta_1 a_1 + \cdots + \beta_k a_k\), then the coefficients are given by \(\beta_i = a_i^T x\)

• Given a set of vectors \(a_1, \ldots, a_k\), the Gram-Schmidt algorithm produces a set of orthonormal vectors \(q_1, \ldots, q_k\) or terminates early if \(q_1, \ldots, q_k\) are linearly dependent. The algorithm is roughly given by for \(i = 1, \ldots, k\)

  1. Orthogonalize: \(\tilde{q}_i = a_i - (q_1^T a_i)q_1 - \cdots - (q_{i-1}^T a_i)q_{i-1}\)
  2. Test for dependence: if \(\tilde{q}_i = 0\), quit
  3. Normalize: \(q_i = \tilde{q}_i / ||\tilde{q}_i||\)