Math4ML

1. **Vector spaces**
    a. Definition
        i. Six properties
        ii. Linear independence
        iii. Span
        iv. Basis
        v. Finite vs Infinite dimensional vector spaces
    b. Euclidean space
        i. Addition and scalar multiplication in R^n
    c. Subspaces
        i. Definition
            1. 0
            2. Closure under addition
            3. Closure under scalar multiplication
        ii. U and W are subspaces, V = U + W is also
        iii. Meaning of **direct sum (**U intersection W = 0**)**
2. **Linear Maps**
    a. Definition
        i. Two fundamental properties that give rise to grid lines being parallel and evenly spaced
    b. Linear operator
    c. Homomorphism of vector spaces
    d. Isomorphism (The inverse is also a homomorphism)
        i. **Note:** Finite dimensional vector spaces of the same dimension are also isomorphic
    e. Matrices
        i. Some properties
        ii. Null space of a linear map (null space is a subspace of domain)
        iii. Range of a linear map (range is subspace of co-domain)
        iv. Column space (range of linear map)
        v. Row space (range of transpose of linear map)
        vi. Dimension of col-space is same as dimension of row-space = Rank of A
3. **Metric Spaces**
    a. Generalized notion of euclidean distance
    b. Triangle inequality
    c. **Motivation:** Allow for limits in mathematical object (convergence)
4. **Normed Spaces**
    a. Generalized notion of length
    b. P-norm
    c. ∞-norm (max $|x_i|$ for 1<= i <=n)
    d. Equivalence of norms in finite dimensional vector spaces

5. **Inner product spaces**
   a. Definition
   b. Inner product induces a norm
   c. An inner product space is also a norm space (and hence also a metric space)
      i. **Note:** If an inner product space is complete with respect to the distance metric induced my its inner product, we say that it is a **Hilbert space**
         1. **Complete** A metric space *M* is called complete (or a Cauchy space) if every Cauchy sequence of points in *M* has a limit that is also in *M* or, alternatively, if every Cauchy sequence in *M* converges in *M*.
   d. Orthogonal and Orthonormal
   e. Pythagorean Theorem
   f. Cauchy-Schwarz inequality
   g. Orthogonal Complements and Projections
      i. Theorem: Every element of an inner product space can be uniquely decomposed in terms of orthogonal complement and projection of a finite-dimensional subspace of the inner product space.
      ii. The orthogonal projection is a well defined function with various properties (refer the book)
      iii. Properties of P_s, the matrix that transforms an element v of vector space V to it's orthogonal project on the subspace S
         1. Linear map
         2. Identity map when restricted to S
         3. range(P_s) = S and null(P_s) = complement space of S
         4. $P_s^2 = P_s$
      iv. **Projection**
         1. Projection is a linear transformation P from a vector space to itself such that $P^2 = P$ i.e. any idempotent linear transform
         2. Among all of the vectors s present in the subspace S, the one which leads to minimum norm of (v-s) is the orthogonal projection of v in the subspace S. Thus, the matrix P_s solves the optimization problem of finding the closest point in S to a given v in V
      v. P_s can be expressed as the sum-of-outer-products of the vectors constituting the orthonormal basis

Let us now consider the specific case where S is a subspace of $\mathbb{R}^n$ with orthonormal basis $\mathbf{u}_1, \dots, \mathbf{u}_m$. Then

$$P_S\mathbf{x} = \sum_{i=1}^{m} \langle \mathbf{x}, \mathbf{u}_i \rangle \mathbf{u}_i = \sum_{i=1}^{m} \mathbf{x}^\top \mathbf{u}_i \mathbf{u}_i = \sum_{i=1}^{m} \mathbf{u}_i \mathbf{u}_i^\top \mathbf{x} = \left(\sum_{i=1}^{m} \mathbf{u}_i \mathbf{u}_i^\top \right) \mathbf{x}$$

so the operator P_S can be expressed as a matrix

$$\mathbf{P}_S = \sum_{i=1}^{m} \mathbf{u}_i \mathbf{u}_i^\top = \mathbf{U}\mathbf{U}^\top$$

where $\mathbf{U}$ has $\mathbf{u}_1, \dots, \mathbf{u}_m$ as its columns. Here we have used the sum-of-outer-products identity.

6. **Eigenthings**
    a. Ax = λx
    b. Basic properties

7. **Trace**
    a. Definition
    b. Properties
        i. **tr(ABCD) = tr(BCDA) = tr(CDAB) = tr(DABC)**
    c. **Note:** Trace of a matrix is equal to sum of its eigen values

8. **Determinant**
    a. Definition
    b. Properties
    c. **Note:** Determinant of matrix is product of its eigen-values

9. **Orthogonal Matrices**
    a. Definition - A matrix is orthogonal is columns are pairwise orthogonal
    b. Orthogonal matrices preserve inner products (and thus 2-norms). These transformation preserve lengths but may reflect or rotate the vector (about origin)

10. **Symmetric Matrices**
    a. $A = A^T$
    b. **Theorem**

Theorem 2. *(Spectral Theorem) If $\mathbf{A} \in \mathbb{R}^{n \times n}$ is symmetric, then there exists an orthonormal basis for $\mathbb{R}^n$ consisting of eigenvectors of $\mathbf{A}$.*

The practical application of this theorem is a particular factorization of symmetric matrices, referred to as the **eigendecomposition** or **spectral decomposition**. Denote the orthonormal basis of eigenvectors $\mathbf{q}_1, \dots, \mathbf{q}_n$ and their eigenvalues $\lambda_1, \dots, \lambda_n$. Let $\mathbf{Q}$ be an orthogonal matrix with $\mathbf{q}_1, \dots, \mathbf{q}_n$ as its columns, and $\mathbf{\Lambda} = \operatorname{diag}(\lambda_1, \dots, \lambda_n)$. Since by definition $\mathbf{A}\mathbf{q}_i = \lambda_i \mathbf{q}_i$ for every i, the following relationship holds:

$$\mathbf{AQ} = \mathbf{Q\Lambda}$$

Right-multiplying by $\mathbf{Q}^\top$, we arrive at the decomposition

$$\mathbf{A} = \mathbf{Q\Lambda Q}^\top$$

    c. **Rayleigh quotients**
        i. Quadratic form : x^TAx
        ii. $R_{\mathbf{A}}(\mathbf{x}) = \frac{\mathbf{x}^\top \mathbf{A}\mathbf{x}}{\mathbf{x}^\top \mathbf{x}}$
        iii. Scale invariance $R_A(x) = R_A(ax)$
        iv. If x is an eigenvector then $R_A(x) = \lambda$
        v. **Variational characterization of eigenvalues:** Rayleigh quotient is bounded by the largest and smallest eigenvalues of A (equality iff x is an eigenvector)

11. Positive (semi-)definite matrices

a. Definition: A symmetric matrix A is called PSD if for every vector x in R^n, x^TAx is greater than equal to 0.
b. A matrix is PSD iff all of its eigenvalues are non-negative (proof from variational characterization)
c. **A^TA is always PSD**
d. A is PSD and $\epsilon > 0$ then $A + \epsilon I$ is positive definite
e. **Note:** Using the above two results, (**$A^TA + \epsilon I$**) is positive definite
f. Geometric intuition
   i. **Level set/ Isocontour:** The set of all inputs such that function applied to those inputs yields a given input. c ∈ dom f : f(x) = c
   ii. **Isocontours of f(x) = xTAx are ellipsoids such that the axes point in the directions of the eigenvectors of A, and the radii of these axes are proportional to the inverse square roots of the eigenvalues**

12. Singular value decomposition

a. Sum of outer products identity

$$\mathbf{A} = \sum_{i=1}^{r} \sigma_i \mathbf{u}_i \mathbf{v}_i^\top$$

b. SVD provides eigendecomposition of A^TA and AA^T

$$\mathbf{A}^\top\mathbf{A} = (\mathbf{U\Sigma V}^\top)^\top\mathbf{U\Sigma V}^\top = \mathbf{V\Sigma}^\top\mathbf{U}^\top\mathbf{U\Sigma V}^\top = \mathbf{V\Sigma}^\top\mathbf{\Sigma V}^\top$$
$$\mathbf{AA}^\top = \mathbf{U\Sigma V}^\top(\mathbf{U\Sigma V}^\top)^\top = \mathbf{U\Sigma V}^\top\mathbf{V\Sigma}^\top\mathbf{U}^\top = \mathbf{U\Sigma\Sigma}^\top\mathbf{U}^\top$$

c. Columns of V (right-singular vectors of A) are eigenvectors of A^TA and columns of U (left-singular vectors of A) are eigenvectors of AA^T