VMLS Cheatsheet[1-9] - meanmachin3

1 Vectors

Vectors

An ordered finite list of numbers.
Block or stacked vectors($a = [b, c, d]$), Subvectors ($a_{r:s} = (a_r, ..., a_s)$), Zero vectors (all elements equal to zero), Unit vectors(($e_i = 1$)), Ones vector(1_n) & Sparsity($nnz(x)$)

Vector addition

Commutative: $a + b = b + a$
Associative: $(a + b) + c = a + (b + c)$
$a + 0 = 0 + a = a$
$a - a = 0$

1.1 Scalar-vector multiplication

$(-2)(1, 9, 6) = (-2, -18, -12)$
Commutative: $\alpha a = a\alpha$
Left-distributive: $(\beta + \gamma)a = \beta a + \gamma a$
Right-distributive: $a(\beta + \gamma) = a\beta + a\gamma$

Linear combinations: $\beta_1 a_1 + ... + \beta_m a_m$
· With Unit vectors: $b = b_1 e_1 + ... + b_n e_n$
· If $\beta_1 + ... + \beta_m = 1$, linear combination is said to be *affine combination*

1.2 Inner product

$a^T b = a_1 b_1 + a_2 b_2 + ... + a_n b_n$ **Properties**:
· *Commutativity*:$a^T b = b^T a$
· *Scalar multiplication Associativity*:
$(\gamma a)^T b = \gamma(a^T b)$
· Vector addition Distributivity:
$(a + b)^T c = a^T c + b^T c$.
General examples:
· Unit vector: $e_i^T a = a_i$
· Sum: $\mathbf{1}^T a = a^1 + ... + a^n$
· Average: $(\mathbf{1}/n)^T a = (a^1 + ... + a^n)/n$
· Sum of squares: $a^T a = a_1^2 + ... + a_n^2$
· Selective sum: If $b_i = 1 or 0$, $b^T a$ is the sum of elements for which $b_i = 1$,
Block vectors
$a^T b = a_1^T b_1 + ... + a_k^T b_k$

1.3 Complexity of vector computations

Space: 8n bytes
Complexity of vector operations: $x^T y = 2n - 1$ flops (n scalar multiplications and $n - 1$ scalar additions)

2 Linear functions

2.1 Linear functions

$f : R^n \to R$ means f is a function mapping n-vectors to numbers
Superposition & linearity: $f(\alpha x + \beta y) = \alpha f(x) + \beta f(y)$
$\cdot f(\alpha_1 x_1 + ... + \alpha_k x_k) = \alpha_1 f(x_1) + ... + \alpha_k f(x_k)$
A function that satisfies superposition is called *linear*
Linear function satisfies
· *Homogeneity*: For any n-vector x and any scalar α, $f(\alpha x) = \alpha f(x)$
· *Additivity*: For any n-vectors x and y, $f(x + y) = f(x) + f(y)$
Affine functions $f : R_n \to R$ is affine if and only if it can be expressed as $f(x) = a^T x + b$ for some n-vector a and scalar b, which is sometimes called the *offset* ·Any *affine* scalar-valued function satisfies the following variation on the super-position property: $f(\alpha x + \beta y) = \alpha f(x) + \beta f(y)$, where $\alpha + \beta = 1$

2.2 Taylor approximation

The (first-order) Taylor approximation of f near (or at) the point z:
$\hat{f}(x) = f(z) + \frac{\partial f}{\partial x_1}(z)(x_1 - z_1) + ... + \frac{\partial f}{\partial x_n}(z)(x_n - z_n)$
Alternatively, $\hat{f}(x) = f(z) + \nabla f(z)^T (x - z)$

2.3 Regression model

Regression model is (the affine function of x) $\hat{y} = x^T \beta + v$

3 Norm and distance

3.1 Norm

Euclidean norm (or just norm) is
$\|x\| = \sqrt{x_1^2 + x_2^2 + ... + x_n^2} = \sqrt{x^T x}$
Properties
· homogeneity: $\|\beta x\| = |\beta| \|x\|$
· triangle inequality: $\|x + y\| \le \|x\| + \|y\|$
· non negativity: $\|x\| \ge 0$
· definiteness: $\|x\| = 0$ only if x = 0
positive definiteness = non negativity + definiteness
$\mathbf{rms}(x) = \sqrt{\frac{x_1^2 + ... + x_n^2}{n}} = \frac{\|x\|}{\sqrt{n}}$
Norm of block vectors $\|(a, b, c)\| = \sqrt{\|a\|^2 + \|b\|^2 + \|c\|^2} = \|(\|a\|, \|b\|, \|c\|)\|$
Chebyshev inequality

3.2 Distance

$\mathbf{dist}(a, b) = \|a - b\|$
Triangle Inequality: $\|a - c\| = \|(a - b) + (b - c)\| \le \|a - b\| + \|b - c\|$
z_j is the nearest neighbor of x if $\|x - z_j\| \le \|x - z_i\|, i = 1, .., m$

3.3 Standard Deviation

de-meaned vector: $\tilde{x} = x - \mathbf{avg}(x)\mathbf{1}$
standard deviation: $\mathbf{std}(x) = \mathbf{rms}(\tilde{x}) = \frac{\|x - (\mathbf{1}^T x/n)\mathbf{1}\|}{\sqrt{n}}$
$\mathbf{rms}(x)^2 = \mathbf{avg}(x)^2 + \mathbf{std}(x)^2$
By Chebyshev inequality, $|x_i - \mathbf{avg}(x)| \ge \alpha \mathbf{std}(x)$ is no more than $1/\alpha^2$ (for $\alpha > 1$)

3.4 Angle

Cauchy–Schwarz inequality: $|a^T b| \le \|a\|\|b\|$

3.5 Complexity

4 Clustering

4.1 Clustering

4.2 A clustering Objective

4.3 The k-means algorithm

4.4 Examples

4.5 Applications

5 Linear Independence

5.1 Linear Independence

5.2 Basis

5.3 Orthonomal Vectors

5.4 Gram–Schmidt algorithm

6 Matrices

6.1 Matrices

6.2 Zero and identity matrices

6.3 Transpose, addition and norm

6.4 Matrix-vector multiplication

6.5 Complexity

7 Matrix examples

7.1 Geometric transformations

7.2 Selectors

7.3 Incidence matrix

7.4 Convolution

8 Linear equations

8.1 Linear and affine functions

8.2 Linear function models

8.3 Systems of linear equations

9 Linear dynamical systems

9.1 Linear dynamical systems

9.2 Population dynamics

9.3 Epidemic dynamics

9.4 Motion of a mass

9.5 Supply chain dynamics