Linear Vector Spaces:

Definition: A linear vector space, $\mathcal{X}$, is a set of elements (vectors) defined over a scalar field, F, that satisfies the following conditions:

1) if $x \in X$ and $y \in X$ then $x+y \in X$. **2)** $x+y=y+x$ **3)** $(x+y)+z=x+(y+z)$

4) There is a unique vector $0 \in X$, such that $x+0=x$ for all $x \in X$.

5) For each vector $x \in X$ there is a unique vector in X, to be called $(-x)$, such that $x+(-x)=0$. **6)** multiplication, for all scalars $a \in F$, and all vectors $x \in X$,

7) For any $x \in X$, $1x=x$ (for scalar 1).

8) For any two scalars $a \in F$ *and* $b \in F$ and any $x \in X$, $a(bx)=(ab)x$.

9) $(a+b)x=ax+bx$. **10)** $a(x+y)=ax+ay$.

Linear Independence:

Consider n vectors $\{x_1, x_2, .., x_n\}$. If there exists n scalars $a_1, a_2, \dots, a_n$, at least one of which is nonzero, such that $a_1x_1+a_2x_2+\dots+a_nx_n=0$, then the $\{x_i\}$ are linearly dependent.

Spanning a Space:

Let X be a linear vector space and let $\{u_1, u_2, \dots, u_n\}$ be a subset of vectors in X. This subset spans X if and only if for every vector $x \in X$ there exist scalars $x_1, x_2, \dots, x_n$ such that $x = x_1u_1+x_2u_2+\dots+x_mu_m$.

Inner Product:

(x, y) for any scalar function of x and y.

1. $(x,y)=(y,x)$ 2. $(x, ay_1+by_2)=a(x,y_1)+b(x,y_2)$
3. $(x,x) \ge 0$, where equality holds iff x is the zero vector.

Norm:

Ascalar function $\|x\|$ is called a norm if it satisfies:

1. $\|x\| \ge 0$ 2. $\|x\|=0$ if and only if $x=0$.
3. $\|ax\|=|a|\|x\|$ 4. $\|x+y\| \le \|x\|+\|y\|$

Angle:

The angle θ bet. 2 vectors x and y is defined by $\cos\theta = \frac{(x,y)}{\|x\|\,\|y\|}$

Orthogonality:

2 vectors $x, y \in X$ are said to be orthogonal if $(x,y)=0$.

Gram Schmidt Orthogonalization:

Assume that we have n independent vectors $y_1, y_2, \dots, y_n$. From these vectors we will obtain n orthogonal vectors $v_1, v_2, \dots, v_n$.

$$v_1 = y_1, \quad v_k = y_k - \sum_{i=1}^{k-1} \frac{(v_i, y_k)}{(v_i, v_i)} v_i,$$

$$where\ \frac{(v_i, y_k)}{(v_i, v_i)} v_i\ is\ the\ projection\ of\ y_k\ on\ v_i$$

Vector Expansions:

$$x = \sum_{i=1}^{n} x_i v_i = x_1v_1 + x_2v_2 + \dots + x_nv_n,$$

$$for\ orthogonal\ vectors, x_j = \frac{(v_j, x)}{(v_j, v_j)}$$

Reciprocal Basis Vectors:

$$(r_i, v_j) = \begin{cases} 0 & i \ne j \\ 1 & i = j \end{cases}, \quad x_j = (r_j, x)$$

To compute the reciprocal basis vectors: set $\mathbf{B}=[\mathrm{v}_1\ \mathrm{v}_2 \dots \mathrm{v}_\mathrm{n}]$,

$\mathbf{R}=[\mathrm{r}_1\ \mathrm{r}_2 \dots \mathrm{r}_\mathrm{n}]$, $\mathbf{R}^T = \mathbf{B}^{-1}$ In matrix form: $\mathbf{x}^v = \mathbf{B}^{-1}\mathbf{x}^s$

Transformations:

A *transformation* consists of three parts:

domain: $X = \{x_i\}$, range: $Y = \{y_i\}$, and a rule relating each $x_i \in X$ to an element $y_i \in Y$.

Linear Transformations: transformation A is *linear* if:

1. for all $x_1, x_2 \in X$, $A(x_1+x_2) = A(x_1) + A(x_2)$
2. for all $x \in X$, $a \in R$, $A(ax) = aA(x)$

Matrix Representations:

Let $\{v_1, v_2, \dots, v_n\}$ be a basis for vector space X, and let $\{u_1, u_2, \dots, u_n\}$ be a basis for vector space Y. Let A be a linear transformation with domain X and range Y: $A(x)=y$

The coefficients of the matrix representation are obtained from

$$A(v_j) = \sum_{i=1}^{m} a_{ij} u_i$$

Change of Basis: $\mathbf{B}_t = [\mathbf{t}_1\ \mathbf{t}_2\ \dots \mathbf{t}_n]$, $\mathbf{B}_w = [\mathbf{w}_1\ \mathbf{w}_2\ \dots \mathbf{w}_n]$

$$\mathbf{A}' = [\mathbf{B}_w^{-1}\mathbf{A}\mathbf{B}_t]$$

Eigenvalues & Eigenvectors: $\mathbf{Az} = \lambda\mathbf{z}$, $|[\mathbf{A} - \lambda\mathbf{I}]| = 0$

Diagonalization: $\mathbf{B} = [\mathbf{z}_1\ \mathbf{z}_2\ \dots \mathbf{z}_n]$,

where $\{\mathbf{z}, \mathbf{z}_2, \dots, \mathbf{z}_n\}$ are the eigenvectors of a square matrix A,

$$[\mathbf{B}^{-1}\mathbf{AB}] = \mathrm{diag}([\lambda_1\ \lambda_2 \dots \lambda_n])$$

Perceptron Architecture:

$\mathbf{a} = hardlim(\mathbf{Wp}+\mathbf{b})$, $\mathbf{W} = [\,{}_1\mathbf{w}^T\ {}_2\mathbf{w}^T\ \dots\ {}_S\mathbf{w}^T]^T$,

$$a_i = hardlim(n_i) = hardlim(\,{}_i\mathbf{w}^T \mathrm{p} + b_i)$$

Decision Boundary:

$${}_i\mathbf{w}^T \mathrm{p} + b_i = 0$$

The decision boundary is always orthogonal to the weight vector. Single-layer perceptrons can only classify linearly separable vectors.

Perceptron Learning Rule

$$\mathbf{W}^{new} = \mathbf{W}^{old} + \mathbf{e}\mathbf{p}^T, \mathbf{b}^{new} = \mathbf{b}^{old} + \mathbf{e},$$
$$where\ \ \mathbf{e} = \mathbf{t} - \mathbf{a}$$

Hebb's Postulate:

"When an axon of cell A is near enough to excite a cell B and repeatedly or persistently takes part in firing it, some growth process or metabolic change takes place in one or both cells such that A's efficiency, as one of the cells firing B, is increased."

Linear Associator: $\mathbf{a} = purelin(\mathbf{Wp})$

The Hebb Rule: *Supervised Form:* $\mathrm{w}_{ij}^{new} = \mathrm{w}_{ij}^{old} + t_{qi}P_{qi}$

$$\mathbf{W} = \boldsymbol{t}_1\boldsymbol{P}_1^T + \boldsymbol{t}_2\boldsymbol{P}_2^T + \dots + \boldsymbol{t}_Q\boldsymbol{P}_Q^T$$

$$\mathbf{W} = [\mathbf{t}_1\ \mathbf{t}_2 \dots \mathbf{t}_Q] \begin{bmatrix} \mathbf{p}_1^T \\ \mathbf{p}_2^T \\ \vdots \\ \mathbf{p}_Q^T \end{bmatrix} = \mathbf{TP}^T$$

Pseudoinverse Rule: $\mathbf{W} = \mathbf{TP}^+$

When the number, R, of rows of $\mathbf{P}$ is greater than the num ber of columns,Q, of $\mathbf{P}$ and the columns of $\mathbf{P}$ are independent, then the pseudoinverse can be computed by $\mathbf{P}^+ = (\mathbf{P}^T\mathrm{P})^{-1}\mathbf{P}^T$

Variations of Hebbian Learning:

Filtered Learning(Ch.14): $\mathbf{W}^{new} = (1-\gamma)\mathbf{W}^{old} + \alpha\mathbf{t}_q\mathbf{p}_q^T$

Delta Rule (Ch.10): $\mathbf{W}^{new} = \mathbf{W}^{old} + \alpha(\mathbf{t}_q - \mathbf{a}_q)\mathbf{p}_q^T$

Unsupervised Hebb (Ch.13): $\mathbf{W}^{new} = \mathbf{W}^{old} + \alpha\mathbf{a}_q\mathbf{p}_q^T$

Taylor: $F(\mathbf{x}) = F(\mathbf{x}^*) + \nabla F(\mathbf{x})^T|_{\mathbf{x}=\mathbf{x}^*}(\mathbf{x}-\mathbf{x}^*) + \frac{1}{2}(\mathbf{x}-\mathbf{x}^*)\nabla^2 F(\mathbf{x})^T|_{\mathbf{x}=\mathbf{x}^*}(\mathbf{x}-\mathbf{x}^*) + \cdots$

Grad $\nabla F(\mathbf{x}) = \left[\frac{\partial}{\partial \mathrm{x}_1}F(\mathbf{x})\ \ \frac{\partial}{\partial \mathrm{x}_2}F(\mathbf{x})\ \dots\ \frac{\partial}{\partial \mathrm{x}_\mathrm{n}}F(\mathbf{x})\right]^T$

Hessian: $\nabla^2 \mathrm{F}(\mathrm{x}) =$

$$\begin{bmatrix} \frac{\partial}{\partial \mathrm{x}_1{}^2}F(\mathbf{x}) & \frac{\partial}{\partial \mathrm{x}_1\,\partial \mathrm{x}_2}F(\mathbf{x}) \dots & \frac{\partial}{\partial \mathrm{x}_1\,\partial \mathrm{x}_\mathrm{n}}F(\mathbf{x}) \\ \frac{\partial}{\partial \mathrm{x}_2\,\partial \mathrm{x}_1}F(\mathbf{x}) & \frac{\partial}{\partial \mathrm{x}_2{}^2}F(\mathbf{x}) \dots & \frac{\partial}{\partial \mathrm{x}_2\,\partial \mathrm{x}_\mathrm{n}}F(\mathbf{x}) \\ \vdots & \vdots & \vdots \\ \frac{\partial}{\partial \mathrm{x}_\mathrm{n}\,\partial \mathrm{x}_1}F(\mathbf{x}) & \frac{\partial}{\partial \mathrm{x}_\mathrm{n}\,\partial \mathrm{x}_2}F(\mathbf{x}) \dots & \frac{\partial}{\partial \mathrm{x}_\mathrm{n}{}^2}F(\mathbf{x}) \end{bmatrix}$$

Directional Derivatives:

1st Dir.Der.: $\frac{\mathbf{p}^T \nabla F(\mathrm{x})}{\|\mathbf{p}\|}$, **2nd Dir.Der.:** $\frac{\mathbf{p}^T \nabla^2 F(\mathrm{x})\mathbf{p}}{\|\mathbf{p}\|^2}$

Minima:

Strong Minimum: if a scalar $\delta > 0$exists, such that $F(x) < F(x+\Delta x)$ for all Δx such that $\delta > \|\Delta x\| > 0$.

Global Minimum: if $F(x) < F(x+\Delta x)$ for all $\Delta x \ne 0$

Weak Minimum: if it is not a strong minimum, and a scalar $\delta > 0$exists, such that $F(x) \le F(x+\Delta x)$ for all Δx such that $\delta > \|\Delta x\| > 0$.

Necessary Conditions for Optimality:

1st-Order Condition: $\nabla \mathrm{F}(\mathbf{x})|_{\mathbf{x}=\mathbf{x}^*} = 0$ (Stationary Points)

2nd-Order Condition: $\nabla^2 \mathrm{F}(\mathbf{x})|_{\mathbf{x}=\mathbf{x}^*} \ge 0$ (Positive Semi-definite Hessian Matrix).

Quadratic fn.: $F(\mathrm{x}) = \frac{1}{2}\mathbf{x}^\mathrm{T}\mathbf{A}\,\mathbf{x} + \mathbf{d}^\mathrm{T}\mathbf{x} + \mathrm{c}$

$$\nabla \mathrm{F}(\mathrm{x}) = \mathbf{Ax} + \mathbf{d},\ \nabla^2 \mathrm{F}(\mathrm{x}) = \mathbf{A},\ \boldsymbol{\lambda}_{min} \le \frac{\mathbf{p}^T\mathbf{A}\,\mathbf{p}}{\|\mathbf{p}\|^2} \le \boldsymbol{\lambda}_{max}$$