1. 矢量运算基本性质

1. 加法及数乘

$\vec{x}+\vec{y}=\vec{y}+\vec{x} \quad \alpha(\vec{x}+\vec{y})=\alpha\vec{x}+\alpha\vec{y} \quad \vec{x}+\vec{y}+\vec{z}=\vec{x}+(\vec{y}+\vec{z}) \quad (\alpha+\beta)\vec{x}=\alpha\vec{x}+\beta\vec{x}$

2. 内积

$$\vec{x}\cdot\vec{y}=\vec{x}^{\top}\vec{y}=\vec{y}^{\top}\vec{x}=\sum_{i=1}^{n}x_iy_i \quad \vec{x},\vec{y}\in R^n$$

3. 欧氏距离：$||\vec{x}-\vec{y}||=\sqrt{\sum_{i=1}^{n}(x_i-y_i)^2}$

4. 线性不相关

若 $\alpha_1\overrightarrow{x_1}+\alpha_2\overrightarrow{x_2}+\alpha_3\overrightarrow{x_3}+\ \cdots\cdots\ +\alpha_n\overrightarrow{x_n}=0$ 仅在 $\alpha_1=\alpha_2=\alpha_3=\ \cdots\cdots\ =\alpha_n=0$ 时成立，则 $\overrightarrow{x_1},\overrightarrow{x_2},\overrightarrow{x_3},\ \cdots\cdots,\overrightarrow{x_n}$ 时不相关

5. 正交：$\vec{x}^{\top}\vec{y}=0$

6. 子空间（U）与矢量（$\vec{x}$）正交：矢量和该子空间中的所有向量正交

立体几何：向量和平面垂直

2. 矩阵的部分性质

1. 对称阵、对角阵、单位阵

- 对称阵：n阶方阵且元素满足 $a_{ij}=a_{ji}$
- 对角阵：n阶方阵且仅有对角线元素不为0
- 单位阵：n阶方阵，且对角线元素均为1，其余元素均为0

2. 特征根

$\mathbf{A}\vec{x}=\lambda\vec{x} \Rightarrow det(\mathbf{A}-\lambda\mathbf{I})=0$ 则该矩阵存在特征根和特征向量

3. 矩阵的秩

$$rank(\mathbf{A})=rank\left(\mathbf{A}^{\top}\right)=rank\left(\mathbf{A}\mathbf{A}^{\top}\right)=rank\left(\mathbf{A}^{\top}\mathbf{A}\right)$$

4. 奇异阵：$det(\mathbf{A})=0 \Rightarrow \mathbf{A}^{-1}$ 不存在

5. 迹（trace）：$tr(\mathbf{A})=\sum_i a_{ii}$ （对角线元素的和） $tr(\mathbf{AB})=tr(\mathbf{BA})$

6. 模：$||\mathbf{A}||=\sqrt{tr\left(\mathbf{A}^{\top}\mathbf{A}\right)}$

7. $(\mathbf{AB})^{\top}=\mathbf{B}^{\top}\mathbf{A}^{\top}$

3. 矢量及矩阵求导

3.1. 标量对矢量求导

1. 梯度

$$\frac{\partial g(\vec{\omega})}{\partial \vec{\omega}} = \begin{bmatrix} \frac{\partial g}{\partial \omega_1} \\ \frac{\partial g}{\partial \omega_2} \\ \vdots \\ \frac{\partial g}{\partial \omega_n} \end{bmatrix} \quad \frac{\partial g}{\partial \vec{\omega}} \in R^n \quad \text{也可记作 } \nabla_{\omega} g$$

2. Hessen 矩阵

$$\frac{\partial^2 g}{\partial \vec{\omega}^2} = \begin{bmatrix} \frac{\partial^2 g}{\partial \omega_1^2} & \frac{\partial^2 g}{\partial \omega_1 \partial \omega_2} & \cdots\cdots & \frac{\partial^2 g}{\partial \omega_1 \partial \omega_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 g}{\partial \omega_n \partial \omega_1} & \frac{\partial^2 g}{\partial \omega_n \partial \omega_1} & \cdots\cdots & \frac{\partial^2 g}{\partial \omega_n^2} \end{bmatrix} \text{也可记作 } \nabla_{\omega}^2 g$$

3. $\frac{\partial f(\vec{w})g(\vec{w})}{\partial \vec{w}} = g(\vec{w})\frac{\partial f(\vec{w})}{\partial \vec{w}} + f(\vec{w})\frac{\partial g(\vec{w})}{\partial \vec{w}}$

4. $\frac{\partial \left(\frac{f(\vec{w})}{g(\vec{w})}\right)}{\partial \vec{w}} = \frac{g(\vec{w})\frac{\partial f(\vec{w})}{\partial \vec{w}} - f(\vec{w})\frac{\partial g(\vec{w})}{\partial \vec{w}}}{g^2(\vec{w})}$

5. $\frac{\partial(\mathbf{A}\vec{g}(\vec{w}))}{\partial \vec{w}} = \mathbf{A}\frac{\partial \vec{g}(\vec{w})}{\partial \vec{w}}$

6. $\frac{\partial \left[\vec{f}(\vec{w})^\top \vec{g}(\vec{w})\right]}{\partial \vec{w}} = \left(\frac{\partial \vec{f}(\vec{w})}{\partial \vec{w}}\right)^\top \vec{g}(\vec{w}) + \left(\frac{\partial \vec{g}(\vec{w})}{\partial \vec{w}}\right)^\top \vec{f}(\vec{w})$

例1：$g(\vec{\omega}) = \vec{\omega}^\top [A]\vec{\omega}$，则可知 $\frac{\partial g(\vec{\omega})}{\partial \vec{\omega}} = \left([A] + [A]^\top\right)\vec{\omega}$

例2：$\frac{\partial \left(\vec{A}^\top \vec{\omega}\right)}{\partial \vec{\omega}} = \left(\vec{A}^\top\right)^\top = \vec{A}$

3.2. 矩阵求导

$$\frac{\partial g([w])}{\partial [w]}=\begin{pmatrix}\frac{\partial g}{\partial w_{11}} & \cdots\cdots & \frac{\partial g}{\partial w_{1n}}\\ \vdots & \vdots & \vdots \\ \frac{\partial g}{\partial w_{m1}} & \cdots\cdots & \frac{\partial g}{\partial w_{mn}}\end{pmatrix} \qquad [w]=\begin{pmatrix} w_{11} & \cdots\cdots & w_{1n}\\ \vdots & \vdots & \vdots \\ w_{m1} & \cdots\cdots & w_{mn}\end{pmatrix}$$

3.3. Taylor 展开

1. $g(w)$ 在 w_0 处展开

$$g(w)=g(w_0)+\left.\frac{\partial g}{\partial w}\right|_{w=w_0}(w-w_0)+\frac{1}{2}\left.\frac{\partial^2 g}{\partial w^2}\right|_{\omega=\omega_0}(w-w_0)^2+\ \cdots\cdots$$

2. $g(\vec{w})$ 在 $\overrightarrow{w_0}$ 处展开 （存疑）

$$g(\vec{w})=g\left(\overrightarrow{w_0}\right)+\left.\left(\frac{\partial g}{\partial \vec{w}}\right)^{\top}\right|_{\vec{w}=\vec{w_0}}\left(\vec{w}-\overrightarrow{w_0}\right)+\frac{1}{2}\left(\vec{w}-\overrightarrow{w_0}\right)^{\top}\left.\frac{\partial^2 g}{\partial \vec{w}^2}\right|_{\omega=\omega_0}\left(\vec{w}-\overrightarrow{w_0}\right)+\ \cdots\cdots$$

3. $g([w])=g([w_0])+tr\left[\left.\left(\frac{\partial g}{\partial [w]}\right)^{\top}\right|_{[w]=[w_0]}([w]-[w_0])\right]+\ \cdots\cdots$

3.4. 无约束条件的优化算法

训练集 $\vec{x}(t)=\left\{\vec{x}^{1},\ \cdots,\vec{x}^{p},\ \cdots,\vec{x}^{N}\right\}$ ，对应教师 $\vec{d}(t)=\left\{d^1,\ \cdots,d^p,\ \cdots,d^N\right\}$

3.4.1. 线性

目标函数 $g(\vec{w})=\frac{1}{2}E\left\{(d(t)-y(t))^2\right\}=\frac{1}{2}E\left\{\left(d(t)-\vec{w}^{\top}\vec{x}(t)\right)^2\right\}$

则可知

$$g(\vec{w})=\frac{1}{2}E\left\{d^2(t)\right\}-E\left\{d(t)(\vec{x}(t))^{\top}\right\}\vec{w}+\frac{1}{2}\vec{w}^{\top}E\left\{\vec{x}(t)(\vec{x}(t))^{\top}\right\}\vec{w}=\frac{1}{2}E\left\{d^2(t)\right\}-\overrightarrow{R_{dx}}^{\top}\vec{w}+\frac{1}{2}\vec{w}^{\top}[R_{xx}]\vec{w}$$

求极值，$\frac{\partial g(\vec{w})}{\partial \vec{w}}=0\ \Rightarrow\ 0-\overrightarrow{R_{dx}}+[R_{xx}]\vec{w}=0\ \Rightarrow\ \vec{w}=[R_{xx}]^{-1}\overrightarrow{R_{dx}}$

3.4.2. 梯度法（非线性，只能求局部极小值）

$g(\vec{w})=\sum_{p=1}^{N}\left(d^p-y^p(\vec{w})\right)^2$ 使得目标函数最小

梯度调整公式：$\Delta\vec{w}=-\eta\frac{\partial g(\vec{w})}{\partial \vec{w}},\eta>0$

将 $g(\vec{w})$ 在 $\vec{w}=\vec{w}(t-1)$ 处 Taylor 展开，

$$g(\vec{w}(t))=g(\vec{w}(t-1))+\left(\frac{\partial g(\vec{w})}{\partial \vec{w}}\right)^{\top}\Bigg|_{\vec{w}=\vec{w}(t-1)}(\vec{w}(t)-\vec{w}(t-1))$$

$$\Rightarrow\ g(\vec{w}(t))-g(\vec{w}(t-1))=\left(\frac{\partial g(\vec{w})}{\partial \vec{w}}\right)^{\top}\Bigg|_{\vec{w}=\vec{w}(t-1)}(\vec{w}(t)-\vec{w}(t-1))$$

要求 $g(\vec{w}(t))<g(\vec{w}(t-1))$ 时，目标函数不断变小，即误差不断的变小

$$\left(\frac{\partial g(\vec{w})}{\partial \vec{w}}\right)^{\top}|_{\vec{w}=\vec{w}(t-1)}<0\ \Leftrightarrow\ \vec{w}(t)-\vec{w}(t-1)>0$$

$$\left(\frac{\partial g(\vec{w})}{\partial \vec{w}}\right)^{\top}|_{\vec{w}=\vec{w}(t-1)}>0\Leftrightarrow\ \vec{w}(t)-\vec{w}(t-1)<0$$

可调整如下，$\Delta\vec{w}=\vec{w}(t)-\vec{w}(t-1)=-\eta\left.\frac{\partial g(\vec{w})}{\partial \vec{w}}\right|_{\vec{w}=\vec{w}(t-1)},\eta>0$

若 $g\left(\vec{w}^{*}\right)=0$ 时存在，$\vec{w}^{*}$ 存在，梯度下降法收敛于局部最小值

3.4.3. 牛顿法

将 $g(\vec{w})$ 在 $\vec{w}(t-1)$ 处 Taylor 展开如下，

$$g(\vec{w}(t))-g(\vec{w}(t-1))=\left.\frac{\partial \vec{g}}{\partial \vec{w}}\right|_{\vec{w}=\vec{w}(t-1)}(\vec{w}(t)-\vec{w}(t-1))+\frac{1}{2}(\vec{w}(t)-\vec{w}(t-1))^{\top}\frac{\partial^2 g}{\partial \vec{w}^2}(\vec{w}(t)-\vec{w}(t-1))$$

令 $\Delta\vec{w}=\vec{w}(t)-\vec{w}(t-1)$，则可知

$$\Delta g(\vec{w})=\left(\frac{\partial g}{\partial \vec{w}}\right)^{\top}\Bigg|_{\vec{w}=\vec{w}(t-1)}\Delta\vec{w}+\frac{1}{2}\Delta\vec{w}^{\top}\mathbf{H}|_{\vec{w}=\vec{w}(t)}\Delta\vec{w}$$

要使得 $\Delta g(\vec{w})$ 极小，则 $\frac{\partial(\Delta g(\vec{w}))}{\partial \vec{w}}=0$，可知

$$\left.\frac{\partial g}{\partial \vec{w}}\right|_{\vec{w}=\vec{w}(t-1)}+\mathbf{H}(t-1)\Delta\vec{w}=0\ \Rightarrow\ \Delta\vec{\omega}=-\mathbf{H}^{-1}\left.\frac{\partial g}{\partial \vec{w}}\right|_{\vec{w}=\vec{w}(t-1)}$$

为防止奇异，则 $\Delta\vec{\omega}=-(\mathbf{H}+\delta)^{-1}\left.\frac{\partial g}{\partial \vec{w}}\right|_{\vec{w}=\vec{w}(t-1)}$

3.4.4. Gauss - Newton 法

目标函数：$g(\vec{\omega})=\frac{1}{2}\sum_{p=1}^{N}\left(d^{p}-y^{p}(\vec{\omega})\right)=\frac{1}{2}\vec{e}(\vec{\omega})^{\top}\vec{e}(\vec{\omega})$

其中 $\vec{e}(\vec{\omega}) = \left[d^1 - y^1(\vec{\omega}) \quad d^2 - y^2(\vec{\omega}) \quad \cdots\cdots \quad d^N - y^N(\vec{\omega})\right]^\top$
将 $\vec{e}(\vec{\omega})$ 在 $\vec{\omega} = \vec{\omega}(t-1)$ 处 Taylor 展开，

$$\vec{e}(\vec{\omega}(t)) = \vec{e}(\vec{\omega}(t-1)) + \left.\frac{\partial \vec{e}(\vec{\omega})}{\partial \vec{\omega}}\right|_{\vec{\omega}=\vec{\omega}(t-1)} (\vec{\omega}(t) - \vec{\omega}(t-1))$$

记 $\mathbf{J}(\vec{\omega}) = \dfrac{\partial \vec{e}(\vec{\omega})}{\partial \vec{\omega}}$，则

$$\mathbf{J}(\vec{\omega}(t-1)) = \left.\frac{\partial \vec{e}(\vec{\omega})}{\partial \vec{\omega}}\right|_{\vec{\omega}=\vec{\omega}(t-1)} \Rightarrow \vec{e}(\vec{\omega}(t)) = \vec{e}(\vec{\omega}(t-1)) + \mathbf{J}(\vec{\omega}(t-1))\Delta\vec{\omega}$$

欲求出 $\vec{\omega}^* = argmin_{\vec{w}^*}(g(\vec{\omega}))$，将上式代入 $g(\vec{\omega})$ 的表达式中可得

$$g(\vec{\omega}) = \frac{1}{2}||\vec{e}(\vec{\omega}(t-1))||^2 + \vec{e}^\top(\vec{w}(t-1))[\mathrm{J}(\vec{w}(t-1))]\Delta\vec{\omega} + \frac{1}{2}(\Delta\vec{\omega})^\top([\mathrm{J}(\vec{\omega}(t-1))]^\top[\mathrm{J}(\vec{\omega}(t-1)])\Delta\vec{\omega}$$

又 $g(\vec{\omega})$ 取极值时，$\dfrac{\partial g(\vec{\omega})}{\partial \vec{\omega}} = 0$，则可知

$$(\vec{e}^\top(\vec{\omega}(t-1))[\mathrm{J}(\vec{\omega}(t-1)])^\top + [\mathrm{J}(\vec{\omega}(t-1)]^\top[\mathrm{J}(\vec{\omega}(t-1)]\Delta\vec{\omega} = 0$$
$$\therefore\ \Delta\vec{\omega} = -[\mathrm{J}(\vec{\omega}(t-1)]^\top[\mathrm{J}(\vec{\omega}(t-1)]^{-1}[\mathrm{J}(\vec{\omega}(t-1)]^\top\vec{e}(\vec{\omega}(t-1))$$

3.4.5. 自然梯度法（黎曼空间）

$$\Delta\vec{\omega} = -\eta\frac{\partial g(\vec{\omega})}{\partial \vec{\omega}}\vec{\omega}^\top\vec{\omega} \qquad \eta > 0$$