Subspaces in Machine Learning

Objectives

- Define subspace
- Establish centrality of subspaces in machine learning
- Introduce dimension of low-rank approximations

"Subspaces" play a key role in machine learning

- Classification/data modeling : $\underline{A}\,\underline{w} = \underline{d}$

$$\underline{A}\,\underline{w} = \left[\underline{a}_1\ \underline{a}_2 \cdots \underline{a}_m\right] \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_m \end{bmatrix} = \sum_{i=1}^{m} \underline{a}_i w_i$$

$\text{Span}\{\underline{a}_1, \underline{a}_2, \cdots \underline{a}_m\}$

- Modeling matrix data $\quad \hat{\underline{R}} = \underline{T}\,\underline{S}$

$$\hat{\underline{R}} = \left[\underline{t}_1\ \underline{t}_2 \cdots \underline{t}_m\right] \begin{bmatrix} \underline{s}_1^T \\ \underline{s}_2^T \\ \vdots \\ \underline{s}_m^T \end{bmatrix} = \sum_{i=1}^{m} \underline{t}_i \underline{s}_i^T \quad \text{or} \quad \hat{\underline{R}} = \left[\hat{\underline{r}}_1\ \hat{\underline{r}}_2 \cdots \hat{\underline{r}}_k\right] \Rightarrow \hat{\underline{r}}_j = \sum_{i=1}^{m} \underline{t}_i s_{ij}$$

Matrix (Taste profile)

Matrix (Scaling)

$\text{Span}\{\underline{t}_1, \underline{t}_2, \cdots, \underline{t}_m\}$

$\text{Span}\{\underline{v}_1, \underline{v}_2, \cdots \underline{v}_m\}$ is a subspace

★ Formally, a **subspace** $S \subseteq \mathbb{R}^N$ (set of N-dim points) satisfies

i) $\underline{0} \in S$ (contains origin)

ii) if $\underline{f}, \underline{g} \in S$, then $\underline{f} + \underline{g} \in S$ (closed under addition)

iii) if $\underline{f} \in S$, then $\alpha \underline{f} \in S$ (closed under scalar mults)

Examples: $S = \{(x,y,z) \mid z = 0\}$ i) $(0,0,0) \in S$ ✓

ii) $(f_x, f_y, 0) + (g_x, g_y, 0) = (f_x + g_x, f_y + g_y, 0) \in S$ ✓

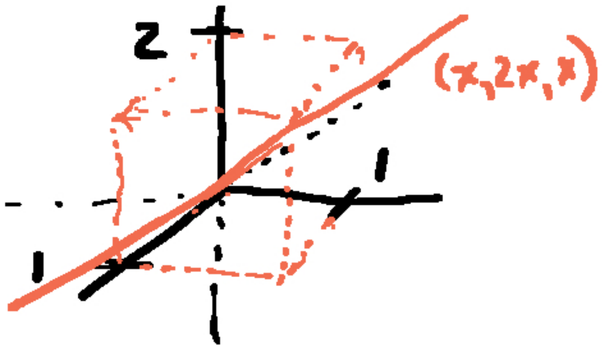
iii) $\alpha(f_x, f_y, 0) = (\alpha f_x, \alpha f_y, 0) \in S$ ✓ Subspace (x-y plane)

- $S = \{(x,y,z) \mid z = 1\}$ i) $(0,0,0) \notin S$ not a subspace

- $S = \{(x, 2x, x)\}$ (line in 3d) i) $(0,0,0) \in S$ ✓

ii) $(f, 2f, f) + (g, 2g, g) = (f+g, 2(f+g), f+g) \in S$ ✓

iii) $\alpha(f, 2f, f) = (\alpha f, 2\alpha f, \alpha f) \in S$ ✓ subspace (line)



Consider $S \subseteq \mathbb{R}^N$, $\left\{ [\underline{v}_1\ \underline{v}_2\ \cdots\ \underline{v}_M] \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_M \end{bmatrix} = \underline{V}\,\underline{w} \text{ for } \underline{w} \in \mathbb{R}^M \right\}$

$\underline{V} : N \times M$

i) if $\underline{w} = \underline{0}$, $\underline{V}\,\underline{w} = \underline{0} \in S$

ii) Let $\underline{f} = \underline{V}\,\underline{w}_f$, $\underline{g} = \underline{V}\,\underline{w}_g$, then $\underline{f} + \underline{g} = \underline{V}\,\underline{w}_f + \underline{V}\,\underline{w}_g = \underline{V}(\underline{w}_f + \underline{w}_g) \in S$

iii) Let $\underline{f} = \underline{V}\,\underline{w}_f$, then $\alpha \underline{f} = \alpha \underline{V}\,\underline{w}_f = \underline{V}(\alpha \underline{w}_f) \in S$

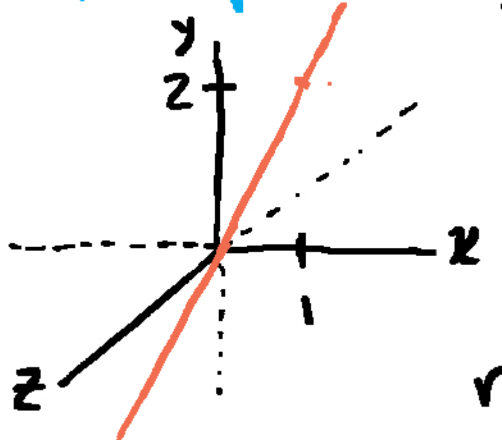
Subspace!

子空间的维度 = V 的 rank (线性独立的向量的最大数量)

☆ **Dimension of $\{\underline{V}\,\underline{w}\}$ is $\text{rank}(\underline{V})$**

Examples:

$\underline{V} = \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}$



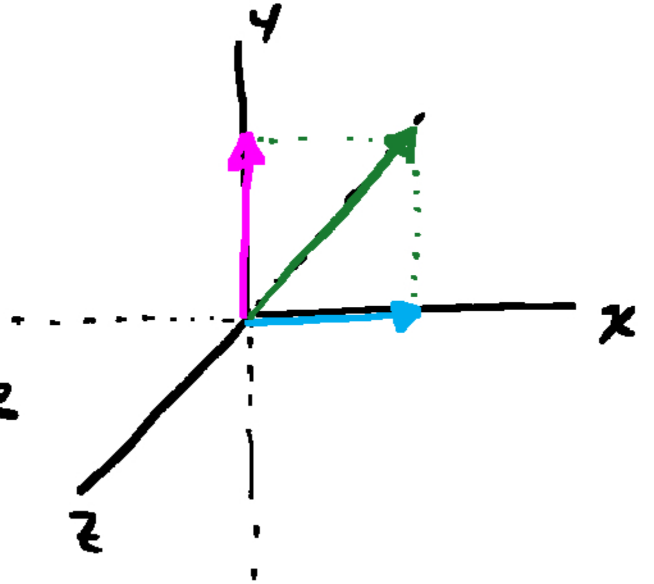
$\underline{V}\,\underline{w}$: line in x-y plane

$\text{rank}(\underline{V}) = 1$

$\underline{V} = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$

$\underline{V}\,\underline{w}$: x-y plane

$\text{rank}(V) = 2$



In general $S = \{\underline{V}\,\underline{w}\}$ is a $K = \text{rank}\,\underline{V}$ dim hyperplane in $\mathbb{R}^N$ (origin)

What about $\underline{\hat{R}} = \underline{T}\,\underline{S}$?

两个 matrix，对子空间的」进一步变换

☆ - In general $\text{rank}(\underline{T}\,\underline{S}) \leq \min\{\text{rank}(\underline{T}), \text{rank}(\underline{S})\}$ (proof in notes)

(rank of multiplied matrices is bounded by the rank of individual matrices)

- Special case: $\underline{T}: N \times M$, $\text{rank}(\underline{T}) = M$ (usually M is $< N$)
  $\underline{S}: M \times K$, $\text{rank}(\underline{S}) = M$ iff

$\text{rank}(\underline{\hat{R}}) = M$ (proof in notes)

This implies $\text{rank}(\bar{R}) = M$

☆ (Matrix 1 has full rank and Matrix 2 has that same rank)

R will be that same rank also?!

$$\underset{N \times K}{\underline{R}} \approx \underset{N \times K}{\underline{\hat{R}}}, \quad \underline{\hat{R}} = \underset{N \times M}{\underline{T}}\ \underset{M \times K}{\underline{S}} = \sum_{i=1}^{M} \underset{\text{columns}}{\underline{t}_i}\, \underline{s}_i^T$$

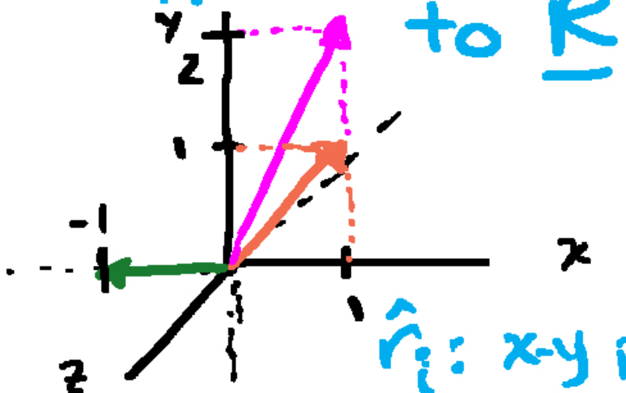
rows of S are weights

rank-M approximation to $\underline{R}$

$$\underline{\hat{R}} = [\underline{\hat{r}}_1, \underline{\hat{r}}_2, \dots, \underline{\hat{r}}_K]$$

$$\underline{\hat{R}} = \underbrace{\begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 0 & 0 \end{bmatrix}}_{\text{rank } 2} \underbrace{\begin{bmatrix} 1 & 1 & -1 \\ 0 & 1 & 1 \end{bmatrix}}_{\text{rank } 2} = \underbrace{\begin{bmatrix} 1 & 1 & -1 \\ 1 & 2 & 0 \\ 0 & 0 & 0 \end{bmatrix}}_{\text{rank } 2}$$

$\underline{\hat{r}}_i$ lie in M-dimensional subspace



$\underline{\hat{r}}_i$: x-y plane

第三部分：核心应用 - 低秩近似

这一部分将前面的数学概念与一个重要的实际应用联系起来。

1. **概念**：
   - $R \approx \hat{R}$，其中 $\hat{R} = TS$
   - 这里的 R 是一个**原始矩阵**，它可能秩很高（或者包含噪声）。
   - $\hat{R}$ 是我们用**低秩模型**去**逼近** R 的结果。
2. **几何解释**：
   - $\hat{R} = [\hat{r}_1, \hat{r}_2, \ldots, \hat{r}_K]$
   - $\hat{R}$ 的每一列 $\hat{r}_i$ 都是一个近似向量。
   - 这些近似向量都是由 T 的列向量线性组合而成的。
   - **关键点**：由于 $\hat{R}$ 的每一列都是由 T 的列向量组合而来，而 T 是一个 $N \times M$ 矩阵（通常 M 远小于 N 和 K），这意味着：
     - **所有**的近似向量 $\hat{r}_i$ 都**位于一个由 T 的列向量所张成的 M 维子空间中**。
     - 这解释了为什么 $\hat{R}$ 的秩是 M。
3. **总结**：
   - 低秩近似的本质就是将原始数据矩阵 R 中的每一个高维向量（N 维）**投影或映射**到一个低维的子空间（M 维）上。
   - 这个投影后的结果就是 $\hat{R}$。

R的维度是 $N \times K$, rank is $\min(N,K)$

↓

近似 matrix $\hat{R}$ $(N \times K)$

由 T 和 S 构建 $\hat{R}$

$T \in N \times M$ (while $M \ll N$)

$S \in M \times K$

$\therefore \text{rank}(\hat{R}) = M$

↓

☆ 降维

Correct answer

Question 5

1 / 1 pts

The span of any three vectors $\mathbf{x}_1$, $\mathbf{x}_2$, and $\mathbf{x}_3$ defines a subspace.

- [x] True
- [ ] False

Quiz Score: **5** out of 5

◂ Previous

Next ▸