Day 13 Training Report

9 July 2025

Cluster Evaluation — Elbow Method, Silhouette Score

On **Day 13**, the focus was on **evaluating the quality of clusters**, as unsupervised learning lacks predefined labels. Two key evaluation methods were introduced:

---

1. Elbow Method

- Used to **determine the optimal number of clusters (k)**.
- Steps:
    1. Compute **Within-Cluster Sum of Squares (WCSS)** for different k values.
    2. Plot WCSS vs k.
    3. Look for the **"elbow point"**, where adding more clusters doesn't significantly reduce WCSS.

Example:

```
import matplotlib.pyplot as plt
wcss = []
for i in range(1, 6):
   kmeans = KMeans(n_clusters=i, random_state=42)
   kmeans.fit(data)
   wcss.append(kmeans.inertia_)

plt.plot(range(1,6), wcss, marker='o')
plt.title('Elbow Method')
plt.xlabel('Number of clusters')
plt.ylabel('WCSS')
plt.show()
```

---

2. Silhouette Score

- Measures **how similar a point is to its own cluster compared to other clusters**.
- Score ranges from -1 to 1:
    - Closer to 1 → well-clustered
    - Closer to 0 → overlapping clusters
    - Negative → wrong clustering

```
from sklearn.metrics import silhouette_score

score = silhouette_score(data[['X','Y']], kmeans.labels_)
print("Silhouette Score:", score)
```