**INFO 204 Practical Test 2 Cheat Sheet**

## **Silhouette Analysis – k-means clustering with (optional) scaling**

perf = []

scaler = StandardScaler();

X\_scale = scaler.fit\_transform(X)

for n\_cluster in [ 2, 3, 4, 5, 6, 7, 8, 9, 10 ]:

km = KMeans(n\_clusters=n\_cluster, random\_state=1234).fit(X)

perf.append({ 'num\_clusters' : n\_cluster, 'score' : silhouette\_score(X, km.labels\_) })

perf = pd.DataFrame(perf)

sns.lineplot(data=perf, x='num\_clusters', y='score')

sns.scatterplot(data=perf, x='num\_clusters', y='score')

plt.show()