Pattern Recognition and Machine Learning: Homework 2

Qingru Hu   2020012996

April 24, 2023

Problem 1

Task1

1.

Take the object function in Equation(2) into three parts:

$$\begin{aligned}
f_1 &= \frac{1}{l}\mathrm{Tr}((Y - JKW)^T(Y - JKW)) \\
f_2 &= \gamma_A \mathrm{Tr}(W^T KW) \\
f_3 &= \frac{\gamma_I}{(u+l)^2}\mathrm{Tr}(W^T KLKW)
\end{aligned}$$

Take the derivative of f_1:

$$\begin{aligned}
\mathrm{d}f_1 &= \frac{1}{l}\mathrm{Tr}(\mathrm{d}(Y - JKW)^T(Y - JKW) + (Y - JKW)^T\mathrm{d}(Y - JKW)) \\
&= \frac{2}{l}\mathrm{Tr}((Y - JKW)^T\mathrm{d}(Y - JKW)) \\
&= \frac{2}{l}\mathrm{Tr}((Y - JKW)^T(-JK)\mathrm{d}W) \\
\frac{\mathrm{d}f_1}{\mathrm{d}W} &= -\frac{2}{l}(JK)^T(Y - JKW)
\end{aligned}$$

Take the derivative of f_2:

$$\begin{aligned}
\mathrm{d}f_2 &= \gamma_A \mathrm{Tr}(\mathrm{d}W^T KW + W^T K\mathrm{d}W) \\
&= 2\gamma_A \mathrm{Tr}(W^T K\mathrm{d}W) \\
\frac{\mathrm{d}f_2}{\mathrm{d}W} &= 2\gamma_A KW
\end{aligned}$$

Take the derivative of f_3:

$$\begin{aligned}
\mathrm{d}f_3 &= \frac{\gamma_I}{(u+l)^2}\mathrm{Tr}((\mathrm{d}W^T)KLKW + W^T KLK\mathrm{d}W) \\
&= \frac{2\gamma_I}{(u+l)^2}\mathrm{Tr}(W^T KLK\mathrm{d}W) \\
\frac{\mathrm{d}f_3}{\mathrm{d}W} &= \frac{2\gamma_I}{(u+l)^2}KLKW
\end{aligned}$$

The solution for the optimization should satisfy:

$$
\begin{aligned}
\frac{\mathrm{d}f_1}{\mathrm{d}W} + \frac{\mathrm{d}f_2}{\mathrm{d}W} + \frac{\mathrm{d}f_3}{\mathrm{d}W} &= 0 \\
-\frac{2}{l}(JK)^T(Y - JKW) + 2\gamma_A KW + \frac{2\gamma_I}{(u+l)^2}KLKW &= 0 \\
((JK)^T JK + \gamma_A lK + \frac{\gamma_I l}{(u+l)^2}KLK)W &= (JK)^T Y \\
\because K^T = K, J^T = J, \text{multiply } K^{-1} & \text{ on both sides to the right} \\
(JJK + \gamma_A lI + \frac{\gamma_I l}{(u+l)^2}LK)W &= JY \\
\because JY &= Y \\
W^* = (JJK + \gamma_A lI + \frac{\gamma_I l}{(u+l)^2}LK)^{-1}Y
\end{aligned}
$$

From the definition of J, we can also have $JJ = J$. Therefore, the best solution is:

$$
W^* = (JK + \gamma_A lI + \frac{\gamma_I l}{(u+l)^2}LK)^{-1}Y
$$

2.

```
# define Y
l = len(X_l)
u = len(X_u)
X = np.concatenate([X_l, X_u])
self.X = X
Y_u = np.zeros([u, Y_l.shape[1]])
Y = np.concatenate([Y_l, Y_u])

self.W = np.linalg.inv(J.dot(K) + self.gamma_A*l*np.identity(l+u)
          + (self.gamma_I*l)/(u+l)**2*L.dot(K)).dot(Y)
```

Task2

The mean of the accuracy of LapRLS is better than that of RLS in two datasets. The std of the two methods are almost the same.

	digits		usps	
	mean	std	mean	std
RLS	0.826	0.0378	0.748	0.0089
LapRLS	0.861	0.0230	0.893	0.0096

Task3

1.

```
from sklearn.decomposition import PCA
from sklearn.discriminant_analysis import LinearDiscriminantAnalysis as LDA
```

```
from sklearn.manifold import MDS
from sklearn.manifold import Isomap
from sklearn.manifold import LocallyLinearEmbedding as LLE
from sklearn.manifold import TSNE

methods = {
  'PCA': PCA(n_components=2),
  'LDA':LDA(),
  'MDS':MDS(n_components=2),
  'Isomap':Isomap(n_components=2),
  'LLE':LLE(n_components=2),
  't-SNE':TSNE(n_components=2)
}
```

2.

Use the digits dataset for visualization. Visualization results are shown in Fig.1 and Fig.2.

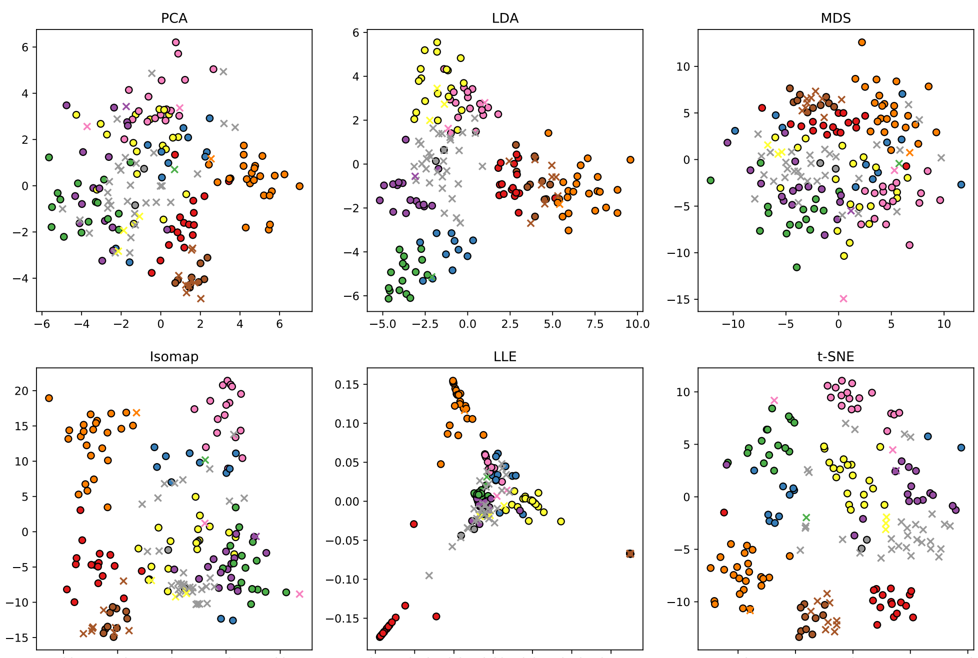

Figure 1: The visualization for RLS

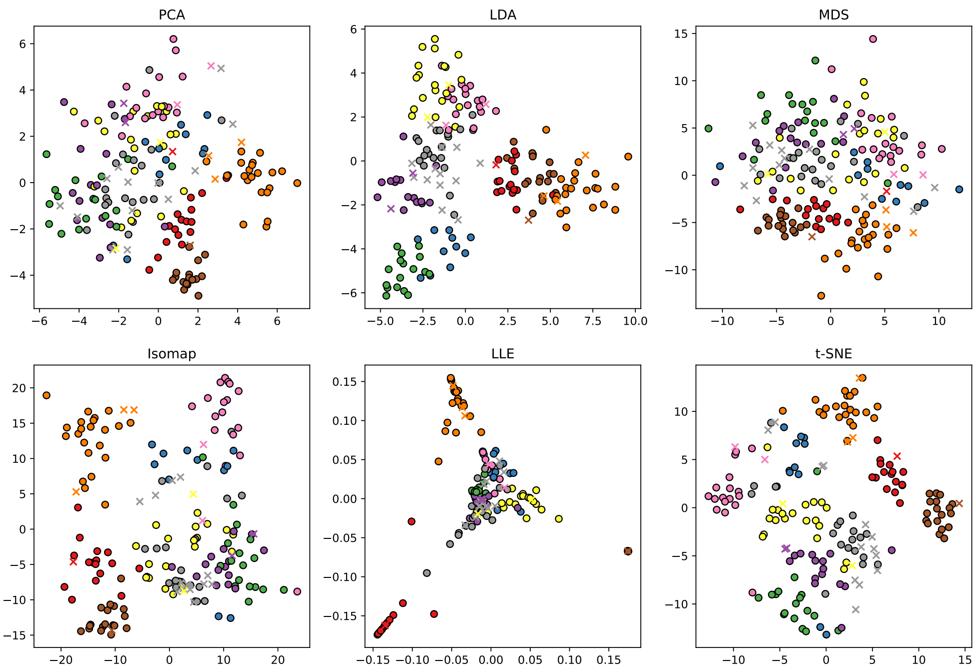

Figure 2: The visualization for LapRLS