DWD Application on Iris Data

Xiaosun Lu

June 24, 2011

- Head of the Data

	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	class
1	5.10	3.50	1.40	0.20	setosa
2	4.90	3.00	1.40	0.20	setosa
3	4.70	3.20	1.30	0.20	setosa
4	4.60	3.10	1.50	0.20	setosa
5	5.00	3.60	1.40	0.20	setosa
6	5.40	3.90	1.70	0.40	setosa

- The Classes

	setosa	versicolor	virginica
Number	50	50	50

View the original data.

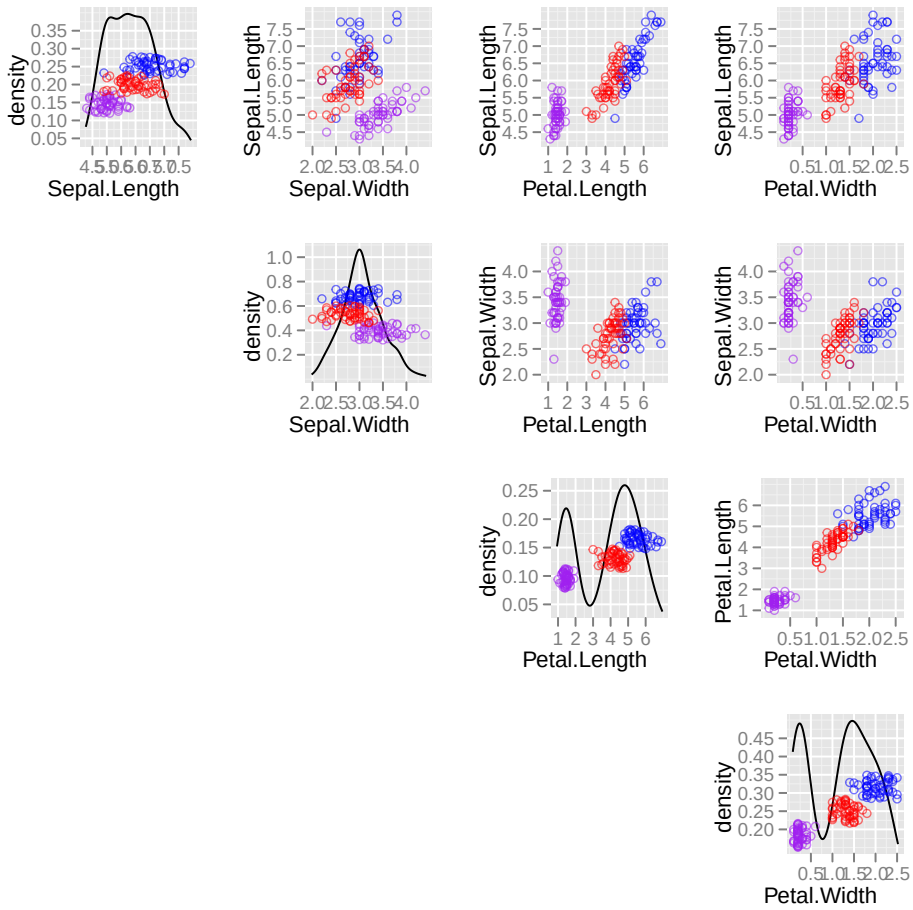

Figure 1: Scatterplots.

Rotate the point cloud onto PC directions.

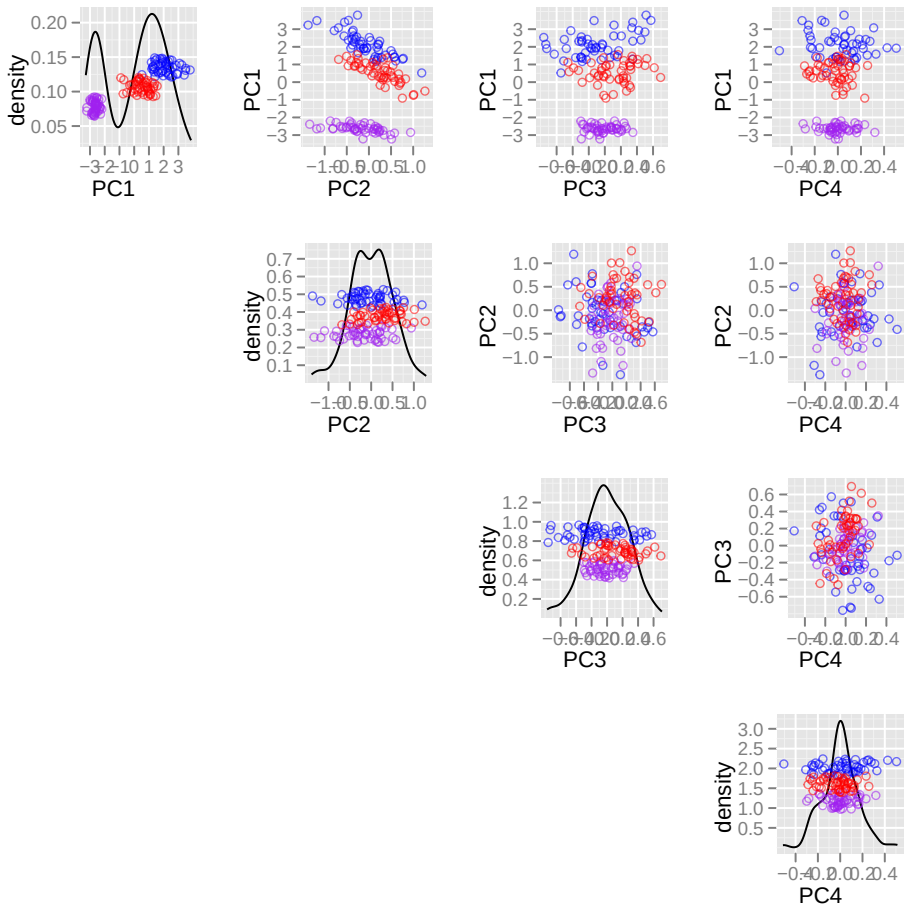

Figure 2: Data projection onto PC directions.

- Among all three species, Versicolor and Virginica are hardest to distinguish.
- Only consider classifying Versicolor Vs. Virginica.
- Hanwen's DWD result is almost identical to Jason's, but much faster.
- The OOB error rate of Random Forest is 0.07 , and the running time is 0.026 s.
- SVM and LDA are also performed on these two classes and then compared with Hanwen's DWD.

- Two-class DWD classification: Versicolor Vs. Virginica
- The angle between the DWD direction by Jason and that by Hanwen is 0.00011382 degree.
- Jason's DWD running time is 74.97 s, and Hanwen's is 3.575 s.

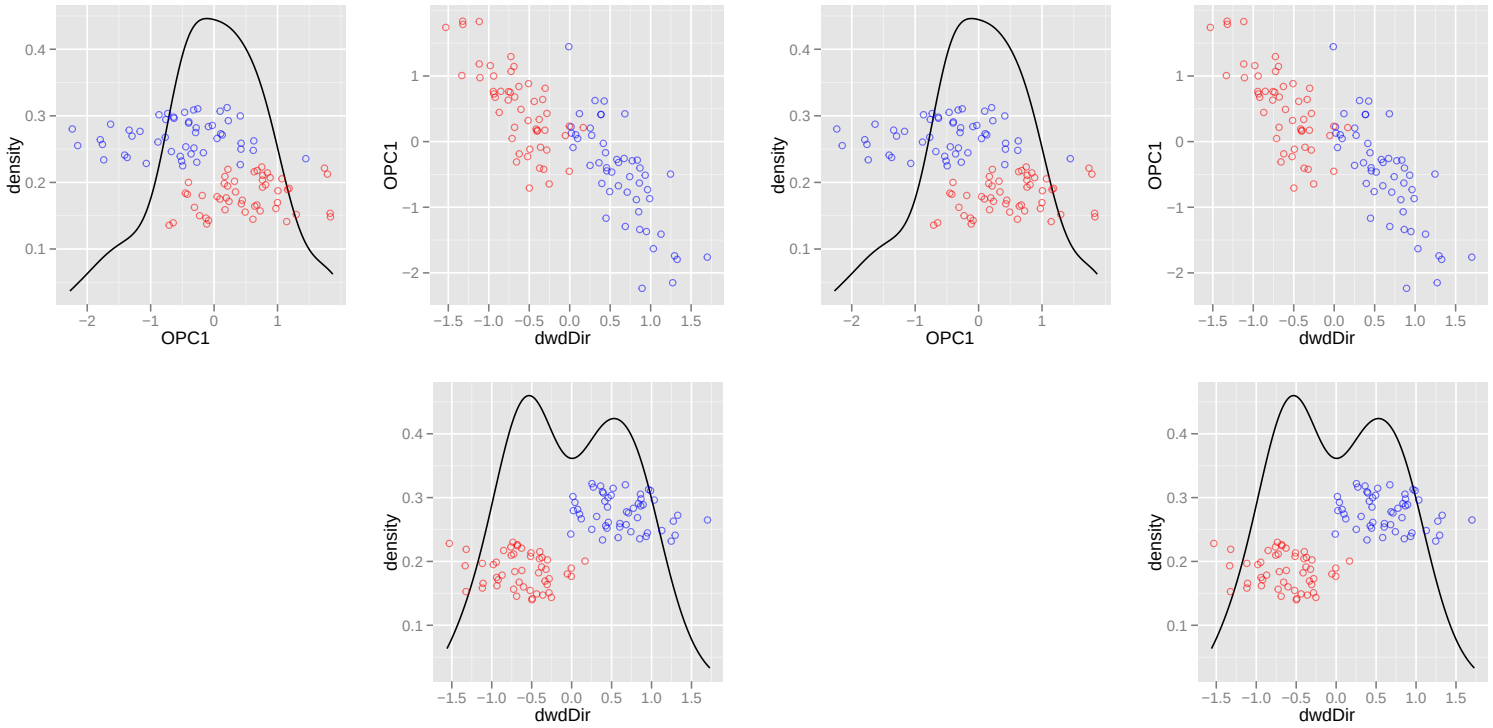

Figure 3: DWD on Iris data for separating versicolor (red) and virginica (blue). The left is Jason's result. The right is Hanwen's result.

- Cross-validation to compare Hanwen's and Jason's DWD (nrep=100, 80% trained)
- Jason's prediction error rate is 0.05 , and the 95% CI is (0, 0.15) .
- Hanwen's prediction error rate is 0.05 , and the 95% CI is (0, 0.15) .
- The angle between Jason's and Hanwen's DWD direction is 0.00009775 degree , and the 95% CI is (0.0000062, 0.00045247)

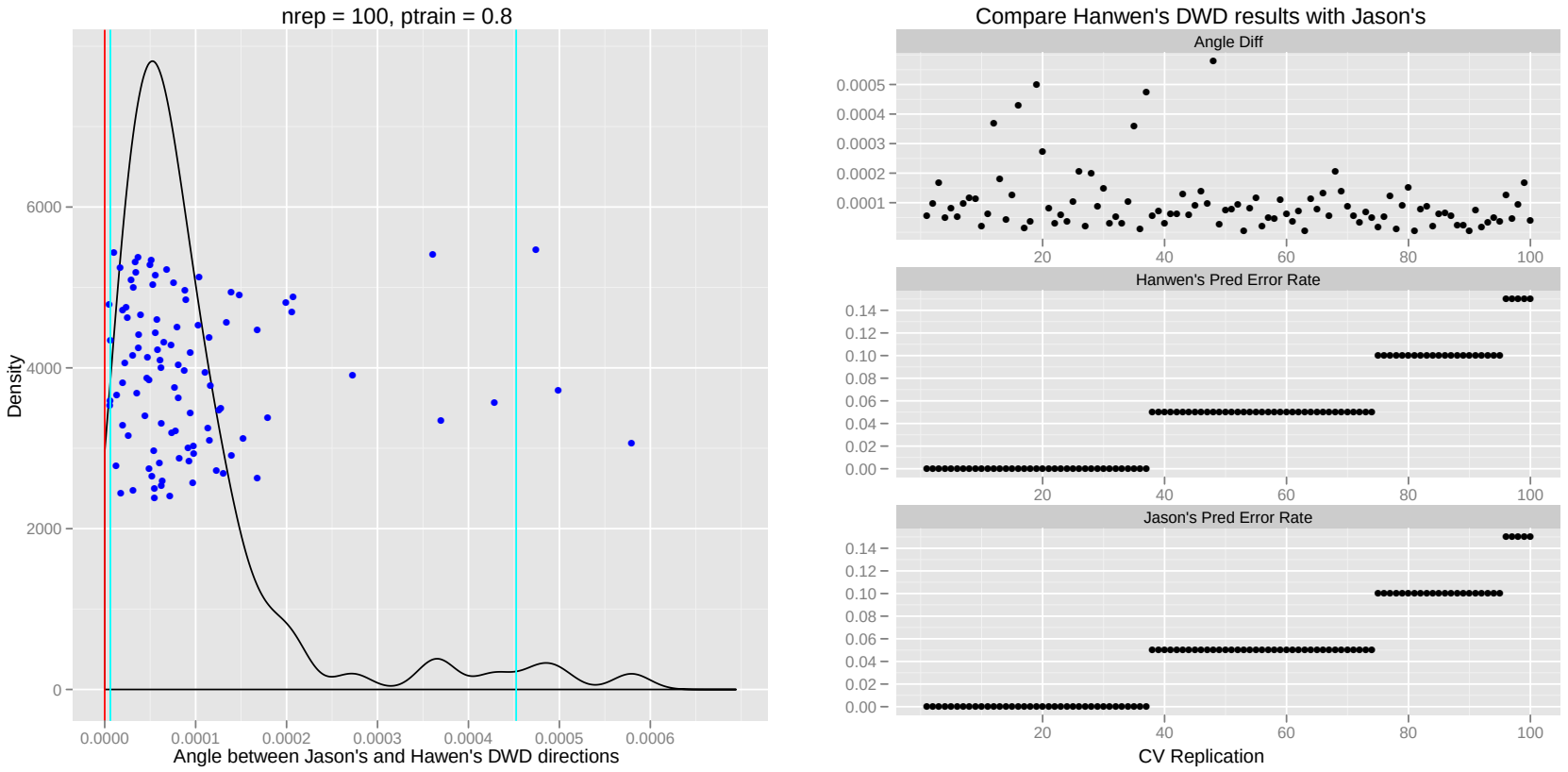

Figure 4: DWD classification. The left is the distribution of angle difference between Jason's and Hanwen's DWD directions. The right shows the angle differences and the prediction errors during the cross-validation.

- Two-class linear SVM classification: Versicolor Vs. Virginica
- SVM running time is 0.015 s, and Hanwen's DWD running time is 3.575 s.
- CV: nrep=100, 80% trained.
- The angle between linear SVM direction and Hanwen's DWD direction is 7.81 degree, and the 95% CI is (2.76, 12.15) degree.
- The linear SVM prediction error rate is 0.03 , and the 95% CI is (0, 0.15)
- Hanwen's DWD prediction error rate is 0.05 , and the 95% CI is (0, 0.15) .

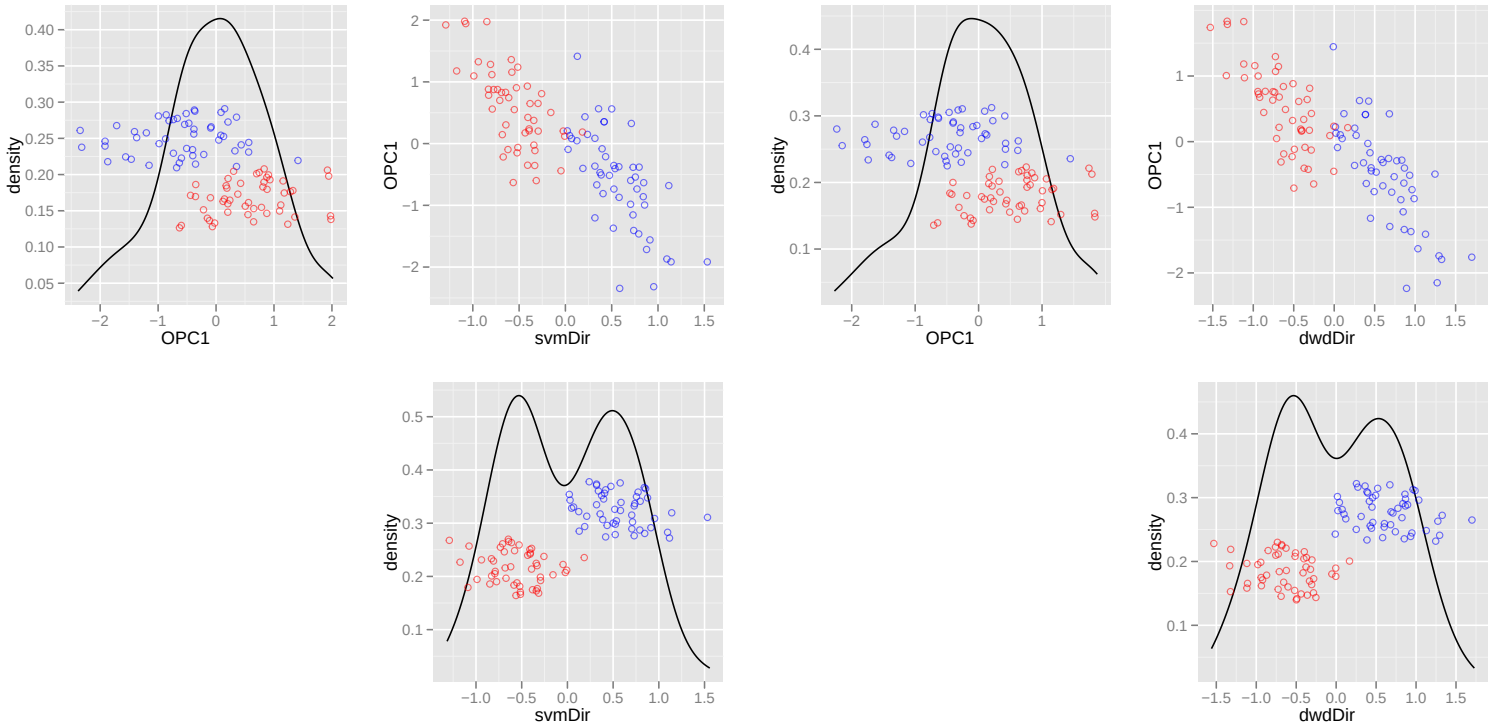

Figure 5: Linear SVM and DWD on Iris data for separating versicolor (red) and virginica (blue). The left is SVM result. The right is Hanwen's DWD result.

- Two-class LDA classification: Versicolor Vs. Virginica
- LDA running time is 0.007 s, and Hanwen's DWD running time is 3.575 s.
- CV: nrep=100, 80% trained.
- The angle between LDA direction and Hanwen's DWD direction is 23.03 degree, and the 95% CI is (17.5, 28.16) degree.
- The LDA prediction error rate is 0.03 , and the 95% CI is (0, 0.13)
- Hanwen's DWD prediction error rate is 0.05 , and the 95% CI is (0, 0.15) .

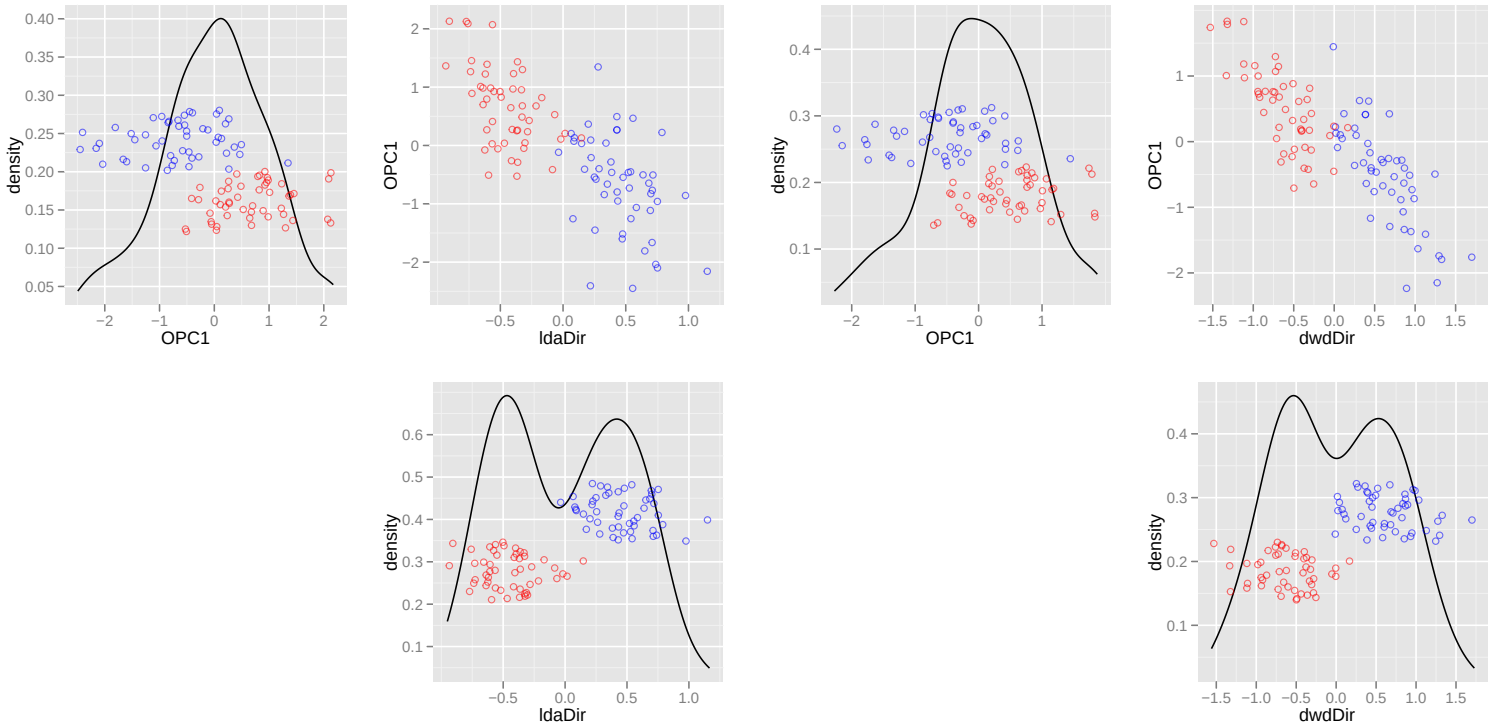

Figure 6: LDA and DWD on Iris data for separating versicolor (red) and virginica (blue). The left is LDA result. The right is Hanwen's DWD result.