Gold Standard

Dear 229 Project TAs,
Please accept our milestone report (attached as a separate pdf) and this brief summary:

<table>
<tr><td>Members</td><td>Dave Deriso, Bryan McCann, Vincent van Gogh</td></tr>
<tr><td>Title</td><td>Classification of Van Gogh's Irises</td></tr>
<tr><td>Predicting</td><td>We aim to classify the kind of irises in Van Gogh's garden</td></tr>
<tr><td>Data</td><td>Data are publicly available here (1 .csv file). There are (150) rows, each representing a unique example flower. There are (5) columns: Sepal Length (continuous), Sepal Width (continuous), Petal Length (continuous), Petal Width (continuous), and Species (class). Species is labeled and represent our ground-truth data with which we will train our model.</td></tr>
<tr><td>Features</td><td>We have a 10 dimensional feature space consisting of: [the first (4) columns of the raw input data, 3 features based on PCA, 3 features based on gaussian kernels]. Our literature review has indicated that a lower dimensional representation of these data will improve the model more than the raw inputs alone (see report). An L1 shrinkage parameter is employed in the regression to discern whether or not a feature is informative (see results section in report).</td></tr>
<tr><td>Models</td><td>We have continuous inputs and have a categorical output which go well with SVM, LDA, multinomial logistic regression, and naive bayes classifiers. We have implemented our models in Python.</td></tr>
<tr><td>Results</td><td>Training set = 75 random samples, Test Set = remaining 75 samples
<table>
<tr><td>Model</td><td>Training MSE</td><td>Test MSE</td><td>N Iterations until Convergence</td></tr>
<tr><td>SVM</td><td>.9432</td><td>.7563</td><td>53</td></tr>
<tr><td>LDA</td><td>...</td><td>...</td><td>...</td></tr>
<tr><td>MNLR</td><td>...</td><td>...</td><td>...</td></tr>
<tr><td>Naive Bayes</td><td>...</td><td>...</td><td>...</td></tr>
<tr><td>Neural Network</td><td>...</td><td>...</td><td>...</td></tr>
</table></td></tr>
<tr><td>Future</td><td>We are planning to continue adding new features, and we may even try some more advanced algorithms. We'd like to tap into some of the deep learning/neural network approaches Professor Ng mentioned, but we're not sure that we have enough data. If this doesn't go well, we are going to focus on improving existing models by training hyperparameters (such as the shrinkage term).</td></tr>
<tr><td>Specific Questions</td><td>We have no specific questions, but will come to office hours for more help. If you must give us feedback, we'd like to know what other methods we should look at, but at this time there are no burning questions.</td></tr>
</table>

Cheers,
Dave Deriso, Bryan McCann, Vincent van Gogh

Details About Each Section

Predicting	What are you building? Specifically, what are the inputs and outputs. Don't explain the motivation for your topic. **(1-2 sentences max)**
Data	Exactly where did your data come from and what does your contain? (ie. What are in the rows and columns? Are examples labeled with ground truth? If you have images, are they color, normalized, etc?) **(2-3 sentences max)**
Features	How many features do you have and which features are the raw input data (ex. color, weight, location, etc) vs. features you have derived (ex. ICA, Gaussian Kernel)? Why they are appropriate for this task? **(3-4 sentences max)**
Models	Exactly which model(s) are you using? Why they are appropriate for this task? Why have you chosen to compare/contrast each model? **(3-4 sentences max)**
Results	Make a compact table of results. Each row should be a different model. The columns should be the training error and the test error. List how many samples are in each of the training and testing data sets. Obviously, these sets should be different. **(1-2 sentences max + 1 table max)**
Future	List plans for the future. **(2-3 sentences max)**
Specific Questions	In brief, if we could provide specific feedback about some aspect of your project, what could it be? Please make this succinct. **(1-2 sentences max)**