Inertial Sensing

EE382V Activity Sensing and Recognition

UT Austin • Dept. Electrical and Computer Engineering • Fall 2016

Today + Admin

Panel of Experts: **Check your grade** on Canvas or **Sign up**

Summary + Critique (HW#3)

Machine Lab (HW#2)

Analyze Inertial Sensor Data

Summary + Critique

HemaApp

> “The interesting question that remains open, is if there is going to be a time in which we have powerful sensors and a strong data base that through machine learning doctors will no longer be required for making diagnosis.”

Summary + Critique

HemaApp

> “one of the purpose of smartphone based application is to provide the health worker in developing country a cheap solution for screening anemia. However, the authors neglect the fact that smartphone is not as popular as in the developed country! Smartphone itself is already expensive in those country. To make the solution more accessible for them, cheaper platform should be chosen. This application can probably be only helpful for people in the developed country.”

Summary + Critique

Simple Sensors

> “Furthermore, they chose the naïve Bayes classifier due to its easy of training and ability to be “real-time”. However, the requirement of “real-time” was not motivated. If we are just looking for long term lifestyle changes, why do we need the system to be real time? Why not choose a more accurate but slower classifier?”

Machine Learning Lab Homework

Q1. *(40 pts)* The dataset in q1_data.csv is comprised of 1000 instances, each with 2 features (F1 and F2) and a binary label (0 or 1). The two features are related in a particularly interesting way.

a) Find out what this relationship looks like by plotting a 2-D graph of the two features. You can use Python's matplotlib for this task or any other plotting tool of your choice. Include the plot in your answer. *(5 pts)*

b) Using scikit-learn, fit a logistic regression model to the dataset and evaluate its performance (accuracy) with 10-fold cross-validation. You should report performance using the accuracy measure, averaged across all cross-validation runs. Include your source code. *(10 pts)*

c) Can you think of a way to improve the performance of the model while still employing the logistic regression algorithm? If so, describe how, include your code and present performance results. (Hint: think about creating new features based on F1 and F2). *(15 pts)*

d) Fit a Random Forest model to the dataset and based on what you find out, discuss why you think it performs better or worse than logistic regression. Include your source code. *(10 pts)*

Machine Learning Lab Homework

Q2. *(40 pts)* Support Vector Machines (SVM) are classifiers that aim to maximize the margin separating classes. In addition to linear classification, SVMs can also be used in non-linear classification by using kernel functions. In this case, instances are mapped to a high-dimensional space and separated using one or more hyperplanes.

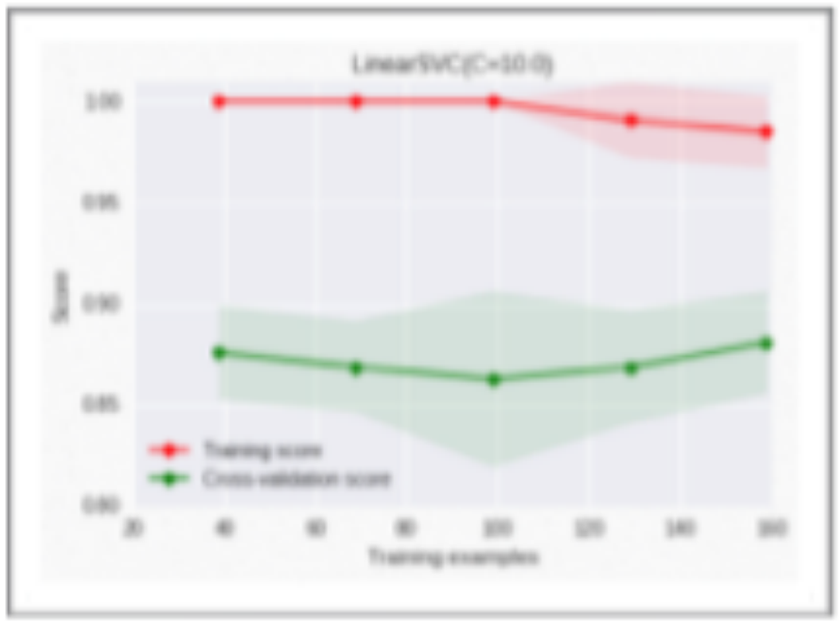
a) Plot the dataset q2_data.csv and indicate whether you think the two dataset classes are linearly separable. Include the plot in your answer. *(5 pts)*

b) Using scikit-learn, fit an SVM model with a **linear** kernel to the dataset and evaluate its performance (accuracy) with 10-fold cross-validation. You should report performance using the accuracy measure, averaged across all cross-validation runs. Include your source code. *(10 pts)*

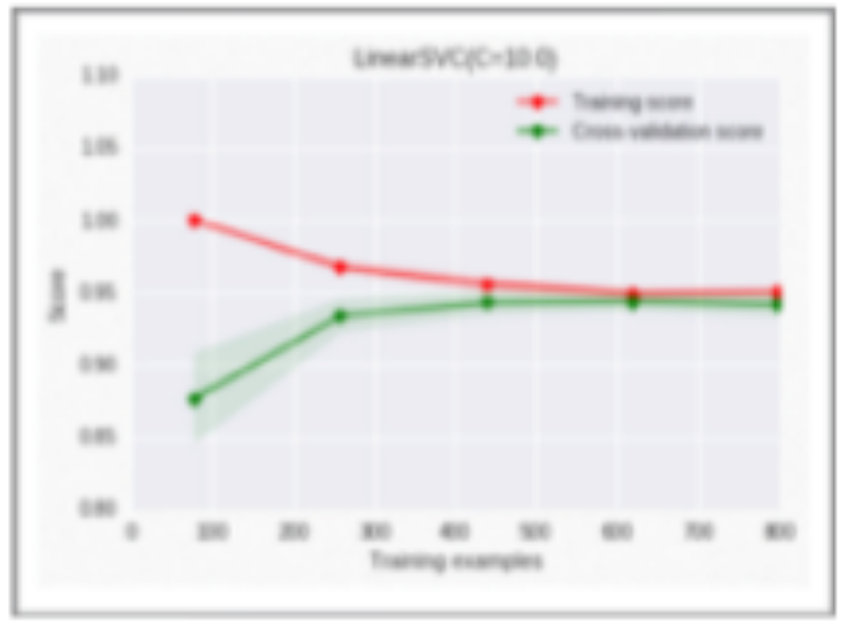
c) Now, fit an SVM model with a **non-linear** kernel to the dataset and evaluate its performance (accuracy) with 10-fold cross-validation. You should report performance using the accuracy measure, averaged across all cross-validation runs. Include your source code. *(10 pts)*

d) Fit a Random Forest model to the q2_data.csv dataset and discuss why you think it performs better or worse than SVM (with linear and non-linear kernel), and how the number of trees in the forest affect the performance results. Include your source code. *(15 pts)*

Machine Learning Lab Homework



Overfitting



Not Overfitting

Provide source code showing one way to reduce overfitting in this example without modifying the classification algorithm (i.e., LinearSVC). (Hint: Think about training data size, regularization and SVM parameterization). *(20 pts)*

Inertial Sensing Lab

Download the file gestures.csv from Canvas

It’s under Files > Inertial Sensing Lab

Load the file within your Python/scikit environment and plot it

Let’s examine this data together…

Inertial Sensing Lab

Now it's your turn!

Change frame and step sizes and see how that affects results

Calculate new features and add them to feature vector

Add labels to the file and build classifier for gestures

Evaluate three classification algorithms with cross-validation

Completing these steps will leave you ready for upcoming homework