Homework 3: Logistic Regression with Regularization

Implement logistic regression as defined in Lecture 10 with feature transformation and regularization for the digits dataset from sklean:

- Transform the feature space to higher dimension with various $\Phi_n(x)$ $(n \geq 5)$. You may call the built-in function in sklearn for polynomial feature transform. (*from sklearn.preprocessing import PolynomialFeatures*)
- You may also perform PCA (Principal Component Analysis) to reduce the feature dimensions. You may call the built-in function available in sklearn for PCA.

Train and validate your implementation with the same dataset for Homework 2 (from sklean):

- Breast cancer, 2 classes, 30 features, 569 data points

To Submit:

Code your work with Python 3. You are supposed to submit both the well-documented .py python files (20pt) and the report. In the report, the following sections are required:

1. **Solution:** (10 pts) Clearly state your algorithm for:
   - Logistic regression with regularization.
   - Computation of the error measure.
   - Feature reduction if applied,
2. **5-fold Cross Validation:** (20 pts)
   - **Experiment:** Description of the setup of the experiments.
   - **Result:** Show the performance plots for various λ (regularization) for both E_{in} and E_{val}. What is the effect of λ? What is the best choice of λ?
   - **Discussion:** Discuss the result: What did you observe? What do you learn from this observation?

Put all files together and submit a zipped file. Include a readme, explaining which problem(s) you have finished. So I know how to grade. Content in the readme file:

1. What did you finish?
2. What python version (2.7? 3.6?)
3. What platform did you use (linux? Mac? windows?)
4. Resources that helped me.