Homework 3: Logistic Regression with Regularization

Implement logistic regression as defined in Lecture 10 with feature transformation and regularization:

- Transform the feature space to higher dimension with various $\Phi_n(x)$: $n = 2, 3$. You may call the built-in function in sklearn for polynomial feature transform. (*from sklearn.preprocessing import PolynomialFeatures*)
- Apply regularization on the problem with input space of $\Phi_n(x)$.
- Train and validate your implementation with the same breast cancer dataset for Homework 2: 2 classes, 30 features, 569 data points
- 5-fold cross validation.

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 with cross validation.
2. **5-fold Cross Validation:** (20 pts)
    - **Experiment:** Description of the setup of the experiments and computation of E_{val}.
    - **Result:** Show the performance plots for various λ (regularization) for E_{val} for various $\Phi_n(x)$, $n = 1, 2, 3$ (similar to slide 17 in lecture 14)
    - **Discussion:** Discuss the result. What did you observe? Does feature transformation to higher order help? What is the effect of λ? What is the best choice of λ? What did you learn from this experiment?

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.