

Programming Ex.8

ML: Programming Exercise 8: Anomaly Detection and Recommender Systems

Debugging Tip

The submit script, for all the programming assignments, does not report the line number and location of the error when it crashes. The following method can be used to make it do so which makes debugging easier.

Open `ex8/lib/submitWithConfiguration.m` and replace line:

```
1 fprintf('!! Please try again later.\n');
2
```

(around 28) with:

```
1 fprintf('Error from file:%s\nFunction:%s\nOn line:%d\n', e.stack(1,1).file,e
    .stack(1,1).name, e.stack(1,1).line );
2
```

That top line says '!! Please try again later' on crash, instead of that, the bottom line will give the location and line number of the error. This change can be applied to all the programming assignments.

error in `ex8_cofi.m` (reported by Charles Davis in session ML-005)

line 199 in `ex8_cofi.m` reads

```
theta = fmincg (@(t)(cofiCostFunc(t, Y, R, num_users, num_movies, num_features, lambda)), initial_parameters, options);
```

but I believe it should be

```
theta = fmincg (@(t)(cofiCostFunc(t, Ynorm, R, num_users, num_movies, num_features, lambda)), initial_parameters, options);
```

...to avoid creating ratings > 5 at line 219. This doesn't affect the submissions of course, just the cosmetics of the recommendations.

Supporting analysis: `Y` is normalized in line 181, creating `Ynorm`, but then it is never used. The video lecture "Implementation Detail: Mean Normalization" at 5:34 makes it pretty clear that the normalized `Y` matrix should be used for calculating `theta`.

This errata also means that "ex8.pdf" Figure 4 is incorrect, since it shows movies with ratings greater than 5-stars.

Item 2: The grader uses `Y` with non-zero values

When using the `R` matrix (to ignore movies that have not been rated), do not rely on `Y(i,j)` to be 0 when a user has not rated a film. This expectation is true for the `ex8_cofi.m` script, but that is NOT true for the test case used by the submit grader for Part 3 through Part 6.

Note: This might no longer be true, the grader seems to be using `Y(i,j) == 0` for when a user has not rated a film

Item 3: Regularization

Note: Unlike previous assignments when we performed regularization, for this exercise, we do NOT skip the 1st column of `Theta` or `X` when computing regularization. This is because we are not specifying bias units in the collaborative filtering algorithm (since the algorithm determines all of the `theta` values, it can set one to the '1' value if it leads to the optimum solution). Therefore, all values of `Theta` and `X` should be considered in regularization.

1.2 Estimating parameters for a Gaussian

the `var` function can actually return normalization with $1/m$ instead of $1/(m-1)$. Set the second argument 0 for $1/(m-1)$ and 1 for $1/m$.

errors in `cofiCostFunc.m`

line 9 should read "% Unfold the **X** and **Theta** matrices from params"

