Errata Week 9

XV. Anomaly Detection

At the risk of being pedantic, it should be noted that p(x) is not a probability but rather the normalized probability density as parameterized by the feature vector, x; therefore, ϵ is a threshold condition on the probability density. Determination of the actual probability would require integration of this density over the appropriate extent of phase space.

In the **Developing and Evaluating an Anomaly Detection System** video an alternative way for some people to split the data is to use the same data for the cv and test sets, therefore the number of anomalous engines (y = 1) in each set would be _20_ rather than 10 as it states on the slide.

XVI. Recommender Systems

In review questions, question 5 in option starting "Recall that the cost function for the content-based recommendation system is" the right side of the formula should be divided by m where m is number of movies. That would mean that the formula will no longer be standard cost function for the content-based recommendation system. However without this change correct answer is marked as incorrect and vice-versa. This description is not very clear but being more specific would mean breaking the honour code.

In the Problem Formulation video the review question states that the no. of movies is $n_m = 1$. The correct value for $n_m = 2$.

In "Collaborative Filtering" video, review question 2: "Which of the following is a correct gradient descent update rule for i ≠ 0?"; Instead of i ≠ 0 it should be k≠0.

In lesson 5 "Vectorization: Low Rank Matrix Factorization" and in lesson 6 "Implementation detail: Mean normalization" the matrix Y contains a mistake. The element $Y^{(5,4)}$ (Dave's opinion on "Sword vs Karate") should be a question mark but is incorrectly given as 0.

In lesson 6 this mistake is propagated to the calculation of μ. When μ is calculated the 5th movie is given an average rating of 1.25 because (0+0+5+0)/4=1.25, but it should be (0+0+5)/3=1.67. This the affects the new values in the matrix Y.

In ex8_cofi.m at line 199, where theta is trained using fmincg() for the movie ratings, the use of "Y" in the function call should be "Ynorm". 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.

In ex8.pdf section 2, "collaborative fitlering" should be "collaborative fi**lt**ering"

In ex8.pdf section 2.2.1, "it will later by called" should be "it will later b**e** called"

In checkCostFunction.m it prints "If your backpropagation implementation is correct...", but in this exercise there is no backpropagation.

In the quiz, question 4 has an invalid phrase: "Even if **you** each user has rated only a small fraction of all of your products (so r(i,j)=0 for the vast majority of (i,j) pairs), you can still build a recommender system by using collaborative filtering." The word "you" seems misplaced, "your" or none.

In the quiz, question 4, one of answer options has a typo "For collaborative filtering, it is possible to use one of the advanced optimization algo**ir**thms"

In ex8.pdf at the bottom of page 8, the text says that the number of features used by ex8_cofi.m is 100. Actually the number of features is 10, not 100.