1 Week 9 Quiz. Anomaly Detection

Anomaly Detection

- Suppose you are developing an anomaly detection system to catch manufacturing defects in airplane engines.
- Your model uses
$$p(x) = \Pi_{j=1}^{n} p(x_j; \mu_j, \sigma_j^2)$$
- You have two features x_1 = ***vibration intensity***, and x_2 = ***heat generated***.
- Both x_1 and x_2 take on values between 0 and 1 (and are strictly greater than 0), and for most "normal" engines you expect that $x_2 \approx x_2$.
- One of the suspected anomalies is that a flawed engine may vibrate very intensely even without generating much heat (large x_1, small x_2), even though the particular values of x_1 and x_2 may not fall outside their typical ranges of values.
- What additional feature x_3 should you create to capture these types of anomalies:

Solution Options

- $x_3 = x_1 + x_2$ This could take on large or small values for both normal and anomalous examples, so it is not a good feature.

Question 1.

For which of the following problems would **anomaly detection** be a suitable algorithm?

- From a large set of hospital patient records, predict which patients have a particular disease (say, the flu).
- From a large set of primary care patient records, identify individuals who might have unusual health conditions.
- CORRECT Given data from credit card transactions, classify each transaction according to type of purchase (for example: food, transportation, clothing).
- In a computer chip fabrication plant, identify microchips that might be defective.

Question 2. Variant A

Suppose you have trained an anomaly detection system that flags anomalies when p(x) is less than ϵ, and you find on the cross-validation set that it has too many false negatives (failing to flag a lot of anomalies). What should you do?

- Increase ϵ [CORRECT]
- Decrease ϵ

Question 2. Variant B

Suppose you have trained an anomaly detection system for fraud detection, and your system that flags anomalies when p(x) is less than ϵ, and you find on the cross-validation set that it mis-flagging far too many good transactions as fradulent. What should you do?

- Increase ϵ
- Decrease ϵ [CORRECT]

Question 3.

Suppose you are developing an anomaly detection system to catch manufacturing defects in airplane engines. You model uses

You have two features x_1 = vibration intensity, and x_2 = heat generated.

Both x1 and x2 take on values between 0 and 1 (and are strictly greater than 0), and for most "normal" engines you expect that $x_1 \approx x_2$.

One of the suspected anomalies is that a flawed engine may vibrate very intensely even without generating much heat (large x1, small x2), even though the particular values of x1 and x2 may not fall outside their typical ranges of values. What additional feature x3 should you create to capture these types of anomalies:

- x3=x1/x2 [CORRECT]
- x3=x21 *times* x2
- x3=x1 *times* x2
- x3=x1+x2

Question 4.

Which of the following are true? Check all that apply.

- When evaluating an anomaly detection algorithm on the cross validation set (containing some positive and some negative examples), classification accuracy is usually a good evaluation metric to use.
- In anomaly detection, we fit a model p(x) to a set of negative (y=0) examples, without using any positive examples we may have collected of previously observed anomalies.
- **CORRECT** When developing an anomaly detection system, it is often useful to select an appropriate numerical performance metric to evaluate the effectiveness of the learning algorithm.
- In a typical anomaly detection setting, we have a large number of anomalous examples, and a relatively small number of normal/non-anomalous examples.

Question 5.

You have a 1-D dataset $\{x(1), \ldots, x(m)\}$ and you want to detect outliers in the dataset. You first plot the dataset and it looks like this:

Suppose you fit the gaussian distribution parameters μ_1 and σ_1^2 to this dataset. Which of the following values for μ_1 and σ_1^2 might you get?

- μ_1=-3,σ_1^2=4
- μ_1=-6,σ_1^2=4
- μ_1=-3,σ_1^2=2
- μ_1=-6,σ_1^2=2

2 Week 9 Quiz. Recommender Systems

Information Filtering System that attempts to recommend information items likely to be of interest to a user.

Commonly used algorithms

- $k-$means clustering
- Pearson's Rho
- Collaborative Filtering

Collaborative Filtering is the process of filtering for information or patterns using collaboration among multiple agents.

Applications: online news aggregation or similar items of clothings

best approached by other methods - prediction

Collaborative Filtering Gradient

$$\frac{\partial J}{\partial X_k^{(i)}} = \sum[]\theta_k^{(j)}$$

$$\frac{\partial J}{\partial \theta_k^{(i)}} = \sum[]X_k^{(j)}$$

No regularization applied

Anomaly Detection Gaussian Distribution Estimate Gaussian Distribution

For n feastures of X , compute the mean and variance for each feature

Selecting Threshold of ϵ.

Implement an algorithm to select the threshold ϵ using an F_i score on a cross validation set.

$P(X) < \epsilon$ is considered to be an anomaly.

2.1 F_1 Score

Recommender Systems

Question 4

Suppose you run a bookstore, and have ratings (1 to 5 stars)
of books. Your collaborative filtering algorithm has learned
a parameter vector $\theta(j) for user j, and a feature$
vector ×(i) for each book. You would like to compute the
"training error", meaning the average squared error of your
system's predictions on all the ratings that you have gotten
from your users. Which of these are correct ways of doing so (check all that apply)?
For this problem, let m be the total number of ratings you
have gotten from your users. (Another way of saying this is
that m=nmi=1nuj=1r(i,j)). [Hint: Two of the four options below are correct.]
1m(i,j):r(i,j)=1((θ(j))T ×(i)r(i,j))2 SELECTED
1m(i,j):r(i,j)=1((θ(j))T ×(i)y(i,j))2
1mnmi=1j:r(i,j)=1(nk=1(θ(j))k ×(i)ky(i,j))2 SELECTED
1mnuj=1i:r(i,j)=1(nk=1(θ(k))j ×(k)iy(i,j))2

Question 2.

In which of the following situations will a collaborative filtering system be the most appropriate learning algorithm (compared to linear or logistic regression)?

- WRONG You're an artist and hand-paint portraits for your clients. Each client gets a different portrait (of themselves) and gives you 1-5 star rating feedback, and each client purchases at most 1 portrait. You'd like to predict what rating your ne×t customer will give you.
- SELECTED You manage an online bookstore and you have the book ratings from many users. You want to learn to predict the e×pected sales volume (number of books sold) as a function of the average rating of a book.
- SELECTED You own a clothing store that sells many styles and brands of jeans. You have collected reviews of the different styles and brands from frequent shoppers, and you want to use these reviews to offer those shoppers discounts on the jeans you think they are most likely to purchase
- WRONG You run an online bookstore and collect the ratings of many users. You want to use this to identify what books are "similar" to each other (i.e., if one user likes a certain book, what are other books that she might also like?)

Question 3

You run a movie empire, and want to build a movie recommendation system based on collaborative filtering. There were three popular review websites (which we'll call A, B and C) which users to go to rate movies, and you have just acquired all three companies that run these websites. You'd like to merge the three companies' datasets together to build a single/unified system. On website A, users rank a movie as having 1 through 5 stars. On website B, users rank on a scale of 1 - 10, and decimal values (e.g., 7.5) are allowed. On website C, the ratings are from 1 to 100. You also have enough information to identify users/movies on one website with users/movies on a different website. Which of the following statements is true?

- You can combine all three training sets into one without any modification and e×pect high performance from a recommendation system.
- It is not possible to combine these websites' data. You must build three separate recommendation systems.
- CORRECT You can merge the three datasets into one, but you should first normalize each dataset separately by subtracting the mean and then dividing by (ma× - min) where the ma× and min (5-1) or (10-1) or (100-1) for the three websites respectively.
- You can combine all three training sets into one as long as your perform mean normalization and feature scaling after you merge the data.

Question 4

Which of the following are true of collaborative filtering systems? Check all that apply.

- WRONG Suppose you are writing a recommender system to predict a user's book preferences. In order to build such a system, you need that user to rate all the other books in your training set.
- CORRECT For collaborative filtering, it is possible to use one of the advanced optimization algoirthms (L-BFGS/conjugate gradient/etc.) to solve for both the ×(i)'s and $\theta(j)$'s simultaneously.
- CORRECT Even if 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.
- WRONG For collaborative filtering, the optimization algorithm you should use is gradient descent. In particular, you cannot use more advanced optimization algorithms (L-BFGS/conjugate gradient/etc.) for collaborative filtering, since you have to solve for both the ×(i)'s and $\theta(j)$'s simultaneously.

Question 5

Suppose you have two matrices A and B, where A is 5×3 and B is 3×5. Their product is C=AB, a 5×5 matri×. Furthermore, you have a 5×5 matri× R where every entry is 0 or 1. You want to find the sum of all elements C(i,j) for which the corresponding R(i,j) is 1, and ignore all elements C(i,j) where R(i,j)=0. One way to do so is the following code:

Which of the following pieces of Octave code will also correctly compute this total? Check all that apply. Assume all options are in code.

- CORRECT total = sum(sum((A * B) .* R))
- CORRECT C = (A * B) .* R; total = sum(C(:));
- WRONG total = sum(sum((A * B) * R));
- WRONG C = (A * B) * R; total = sum(C(:));