Quiz - Classification

1 point

1. The simple threshold classifier for sentiment analysis described in the video (*check all that apply*):

- [x] **Must have pre-defined positive and negative attributes**
- [ ] **Must either count attributes equally or pre-define weights on attributes**
- [ ] **Defines a possibly non-linear decision boundary**

---

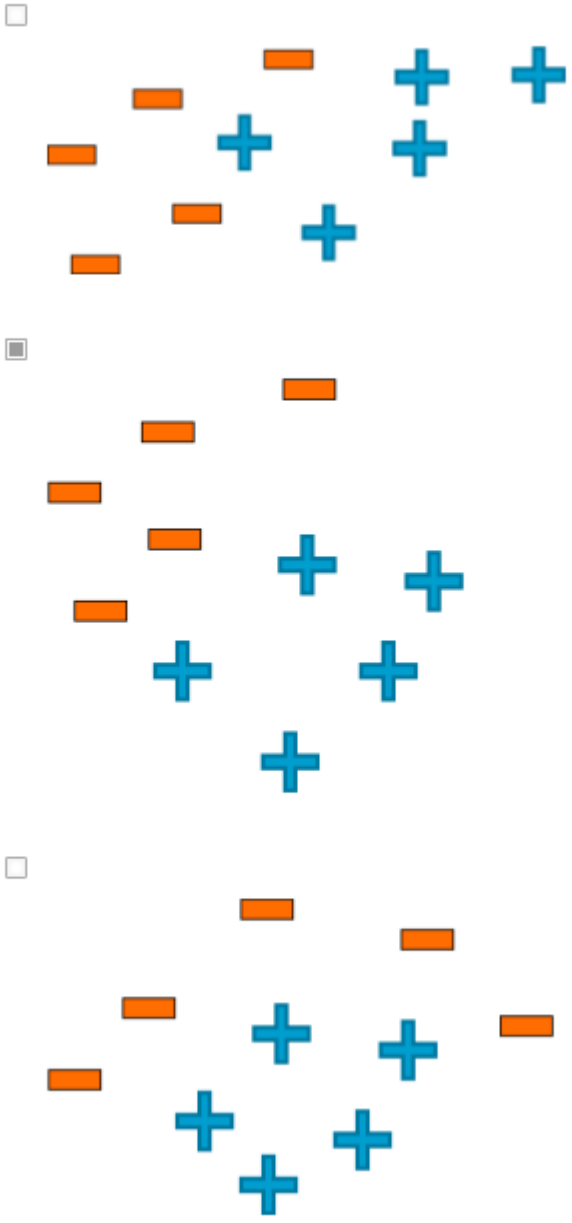
1 point

2. For a linear classifier classifying between "positive" and "negative" sentiment in a review x, Score(x) = 0 implies (*check all that apply*):

- [ ] **The review is very clearly "negative"**
- [x] **We are uncertain whether the review is "positive" or "negative"**
- [ ] **We need to retrain our classifier because an error has occurred**

1 point

3. For which of the following datasets would a **linear** classifier perform perfectly?



1 point

4. ***True or false:*** High classification accuracy always indicates a good classifier.

- ○ **True**
- ◉ **False**

1 point

5. ***True or false:*** For a classifier classifying between 5 classes, there always exists a classifier with accuracy greater than 0.18.

- ◉ **True**
- ○ **False**

1 point

6. ***True or false:*** A false negative is always worse than a false positive.

- ○ **True**
- ◉ **False**

1 point

7. Which of the following statements are true? (*Check all that apply*)

- ☑ **Test error tends to decrease with more training data until a point, and then does not change (i.e., curve flattens out)**
- ☐ **Test error always goes to 0 with an unboundedly large training dataset**
- ☐ **Test error is never a function of the amount of training data**

Assignment - Analyzing Product Sentiment

Analyzing product sentiment

Quiz, 11 questions

1 point

1. Out of the 11 words in *selected_words*, which one is most used in the reviews in the dataset?

- [ ] **awesome**
- [ ] **love**
- [ ] **hate**
- [ ] **bad**
- [x] **great**

1 point

2. Out of the 11 words in *selected_words*, which one is least used in the reviews in the dataset?

- [x] **wow**
- [ ] **amazing**
- [ ] **terrible**
- [ ] **awful**
- [ ] **love**

1 point

3. Out of the 11 words in *selected_words*, which one got the most positive weight in the *selected_words_model*?

(Tip: when printing the list of coefficients, make sure to use print_rows(rows=12) to print ALL coefficients.)

- [ ] **amazing**
- [ ] **awesome**
- [x] **love**
- [ ] **fantastic**
- [ ] **terrible**

1 point

4. Out of the 11 words in *selected_words*, which one got the most negative weight in the *selected_words_model*?

(Tip: when printing the list of coefficients, make sure to use print_rows(rows=12) to print ALL coefficients.)

- ○ **horrible**
- ◉ **terrible**
- ○ **awful**
- ○ **hate**
- ○ **love**

1 point

5. Which of the following ranges contains the accuracy of the *selected_words_model* on the *test_data*?

- ○ **0.811 to 0.841**
- ◉ **0.841 to 0.871**
- ○ **0.871 to 0.901**
- ○ **0.901 to 0.931**

1 point

6. Which of the following ranges contains the accuracy of the *sentiment_model* in the IPython Notebook from lecture on the *test_data*?

- ○ **0.811 to 0.841**
- ○ **0.841 to 0.871**
- ○ **0.871 to 0.901**
- ◉ **0.901 to 0.931**

1 point

7. Which of the following ranges contains the accuracy of the majority class classifier, which simply predicts the majority class on the *test_data?*

- (●) **0.811 to 0.843**
- ( ) **0.843 to 0.871**
- ( ) **0.871 to 0.901**
- ( ) **0.901 to 0.931**

---

1 point

8. How do you compare the different learned models with the baseline approach where we are just predicting the majority class?

- ( ) **They all performed about the same.**
- ( ) **The model learned using all words performed *much better* than the one using the only the *selected_words*. And, the model learned using the *selected_words* performed much better than just predicting the majority class.**
- (●) **The model learned using all words performed much better than the other two. The other two approaches performed about the same.**
- ( ) **Predicting the simply majority class performed much better than the other two models.**

---

1 point

9. Which of the following ranges contains the *'predicted_sentiment'* for the most positive review for *'Baby Trend Diaper Champ'*, according to the *sentiment_model* from the IPython Notebook from lecture?

- ( ) **Below 0.7**
- ( ) **0.7 to 0.8**
- ( ) **0.8 to 0.9**
- (●) **0.9 to 1.0**

---

1 point

10. Consider the most positive review for *'Baby Trend Diaper Champ'* according to the *sentiment_model* from the IPython Notebook from lecture. Which of the following ranges contains the predicted_sentiment for this review, if we use the *selected_words_model* to analyze it?

- ○ **Below 0.7**
- ◉ **0.7 to 0.8**
- ○ **0.8 to 0.9**
- ○ **0.9 to 1.0**

---

1 point

11. Why is the value of the *predicted_sentiment* for the most positive review found using the *sentiment_model* much more positive than the value predicted using the *selected_words_model*?

- ○ **The *sentiment_model* is just too positive about everything.**
- ○ **The *selected_words_model* is just too negative about everything.**
- ○ **This review was positive, but used too many of the negative words in *selected_words*.**
- ◉ **None of the *selected_words* appeared in the text of this review.**