**Things for abstract and Intro**

Abstract

Talk about DLS, lack of specific circumstances (only uses averages), briefly mention the data, machine learning stuff, theory that using the data in more specific circumstances will make it more accurate.

Will need to briefly mention results (so far using 30% of dataset, gets 80-90% accuracy, using 50% gets consistent 93 ish (sometimes rounds to 94))

Briefly explain the overall goal of my thing (predict runs in interrupted overs to replace DLS)

Intro

Go into more detail about cricket, rules explanation and in-depth into D/L to talk about its issues and why I’m trying to make it better. Important distinction between D/L and DLS, but my model by definition will take higher scoring into account

Outline the structure of the diss (lit review, looking at my data, potential dimensionality reduction(not required), GPClassification and sparse)

(Lit review goes here)

Main Body

Hypothesis that using a more complex ML model will not only get a better accuracy on a more difficult prediction than previous ML in sport, but also improve upon the system already in place for cricket (DLS). Maybe to test, see if I can find a game that used DLS in a dodgy way, and use my model to try and predict the interrupted overs, looking intuitively at whether it would make more sense as a target (Instead, to see more concretely whether it works, predict full overs that have already happened to see how close it is). Mine makes use of specific data instead of just averages therefore should be more accurate to the situation, allowing us to predict, ball by ball, how many runs are scored.

Talk more about my data and justify my use of model using GP papers

Possibly design experiments for the DLS stuff, maybe finding a T20 game that has players that have data. Train on different amounts of the dataset, so I can talk about speed vs efficacy of GPC.

Talk about how I would track the runs during interrupted overs (can’t use pen and paper as the data for the players would need to be updated as I go, or not)

Give confidence values/uncertainty (easily done with the sklearn model I believe)

Experimental Results

Show that I run the model on different subsets of the data, it is shuffled to give a more accurate representation of the full dataset, talk about runtime vs accuracy, try to plot results to make it easier to explain. Highlight expected results (I reckon based on my 30% run, expect a bit better with 70-75% but much more runtime as it is O(n^3)). Could refer to appendices for more clear representation

Conclusions

Talk about the results my model achieved (maybe talk about using different sizes of the dataset and about time vs effectiveness) 30% took 45 minutes with 80-90% acc, will be able to measure later for full dataset, compare to other works done in sport (so far does a decent amount better lol)

Compare to DLS (cannot do direct comparison as to which does better but can say that if it predicts the runs with this much accuracy and it takes the specific player stats into account, then it should be more accurate measure than DLS)

Limitations (couldn’t find a library to do sparse GPClassification so done normal, which is O(n^3) whereas sparse is O(m^3)?? Where m<<n, use sparse paper or scalable GP paper stuff)

In future, make it track the predictions (I don’t think that would be very difficult) and try and actually simulate the overs that are interrupted, update the dataframes as each ball comes, could hard code this as we go, by saving the model parameters and inputting each row as we need it, then it would be easier to do in real time, instead of having to train the model fully and then predict each time (obviously unfeasible in any practical sense)