**Rationale**

We would like to predict the character (C) based on the user’s touch point (xy) and the motion data. We start with P(C| xy, motion) and apply Bayes rule, we can get:

(model 1)

Or

(model 2)

If we go with the first model, then the predicted character can be expressed as:

If we go with the second model, then the predicated character can be expressed as:

**Description for each module**

* main\_test.py

This module evaluates the text entry accuracy before and after applying the models. You can choose whatever models you would like to add to the final model.

* touch\_model\_train.py - corresponds to P(xy|C)

This module trains a touch model for each character from the user’s touch points. All the touch points for a character is modeled as an elliptically shaped 2D gaussian distribution. Mathematically, it can be expressed as: . We can interpret this model as “the probability of a touch point given the intended character”. The trained model is saved as *TM.pickle*.

* Bayes\_regression\_model.py – corresponds to P(xy|motion,C)

This module trains a bayes regression model for each character and the phone’s motion data, for example, accelerometer data and gyroscope data. Mathematically, this model can be expressed as:. We can interpret this model as the “probability of a touch point given the intended character and the phone motion”. The trained model is saved as *BLR\_M.pickle*.

* language\_model\_train.py – corresponds to P(C)

This module trains a 5-gram language model with kneser-ney smoothing. It predicts the next character based on the past 4 characters. The trained model is saved as *KNLM.pickle*.

* motion\_model\_train.py – corresponds to P(motion|C)

Similar to the touch model. This module trains a motion model for each character from the phone’s motion. All the motion data for a character is modeled as an elliptically shaped n-dimensional gaussian distribution (for example, n is 3 if motion = [gyro\_x, gyro\_y, gyro z]). However**, it is worth noting that modelling P(motion|C) as a n-d gaussian might not be appropriate. In the experiment, we found that adding this component to the Bayes model will only make the performance worse**. You are advised to either discard this model or use a more appropriate model to model P(motion|C)