Topics that we have discussed about:

* Books to read
* Policy net for Gaussian Distribution
* Future plans in the next eight months

**Books to read:**

There are a few books that I could read or have to complete:

* Machine Learning by Tom Mitchell
* Elements of Statistical Learning
* Learning from Data, a Short Course
* Dynamic Programming and Optimal Control ed 4th vol2
* Markov Decision Processes: Discrete Stochastic Dynamic Programming

Read “Learning from Data, a Short Course” first. Especially, pay attention to Chapter 1.3~2.3, as well as Appendix, which containing the key points in AI research field, which are the trade-off in learning and approximation vs generalization.

**Policy Net for Gaussian Distribution:**

* Distribution models are based on Central limit theorem, and can be used to generalize continuous data
* For the choosing the actions stochastically, we need a distribution model that can deal with continuous data to do the sampling for this case.

**3 Ways I can build the construct the distribution model using Policy Net**

* Dynamic variance: use policy net to compute the mean, and then apply a function like sigmoid function to the mean value to obtain the variance
* Fixed variance: manually testing out and picking a reasonable variance value. Probably can apply some penalties to the fixed variance when it comes up with an unreasonable loss value, but not limited to penalize it in this way.
* No variance, only select mean: doing this way would lead the stochastical action selections to be deterministic action selections

Refer to <https://github.com/chainer/chainerrl/blob/master/chainerrl/policies/gaussian_policy.py>

First two needs to avoid picking an overly small variance, which would lead to deterministic actions.

Need to discover the pros and cons of stochastic actions and deterministic actions from the practice. As well, need to check if fixed variance can outperform dynamic variance, and which one is more convenient to use.

We can do some easy mathematical manipulations to control the range of mean based on the characteristics of tanh(ranges from -1 to 1), and do shifting operations to pick an appropriate mean. Doing this way can guarantee the sampled value to be quite often within the range, but not always.

**KL divergence:**

Computing the KL divergence could be heuristic.

The following is just my thoughts, haven’t search it up yet.

To compute the KL(a | s), we could compare the divergence of the probability P\_theta\_old(a – eps, a + eps ) with P\_theta(a – eps, a + eps )

To compute the KL(\*|s), we pick a set of action values along the available range, and use the scheme above to compute the KL divergence of each one, then take the average of it.

**Future plans for next 8 months:**

1. Investigate the following field:

* Active SLAM
* Interpretable RL, with causality inference
* RL with attention
* Transfer learning in R

Gather information of each field from various resources. Particularly need to have a basic understanding about the challenges of the area that is facing and how could the research on these fields can make benefits in different magnitudes.

2. After picking the favorite topic, need to revise a 3-months plan to study it, particularly need to have a plan about what papers to read.

3. Spend 3 months to study these topics, and make a paper review, also come up with the inspiration from the studying of these paper

4. 3 months of research. (Not quite sure)

**My thoughts and timelines on the plans:**

1. Finish the TRPO humanoid swimmer code first (might takes up to 2 to 3 weeks)

2. Alongside coding up the TRPO, spend time getting to know those four fields. Try to get a broad understanding of each, and pick my favorite one, and revise the plan. (might takes up to 3 weeks, I wish to finish it by the early of February)

3. I am not sure the 3-months for making a paper review you mean is a very productive 3-months, but based on the timeline of my summer break, I would have to finish it around May. Since I am working at Ericson 8 hours per week-day, I am a little bit concerned that if the 3-months is kind of too rush for me.

4. If everything goes well, then that’s the final step (research), then, I would have a plenty of time to do it.