Hi Huanchen and Yetian,

I'm your main TA for the ML project.

This is an interesting problem and your current plan looks good to me.
A few things to keep in mind as you start working on your project:
- Do you already have access to Amazon EC2 for this class project?
- You mentioned that Petuum is still not a mature system. Have you made sure that what you plan to do (L1 and L2 with stochastic (sub)gradient descent) can be done using the current version of Petuum?
- How much of your project involves configuring or modifying Petuum source code? You want to make sure that you have some results by the end of the semester, so please have some backup plans if this turns out to be too time consuming.
- You might want to start with smaller datasets to make sure your implementation runs smoothly, but if you want a really large text dataset there is a Wikipedia dataset that has almost 3 million instances. Let me know if you get to this point.

If you have any questions or would like to discuss ideas, we can schedule a time to meet up. If you prefer to talk over email, that's fine too.

Good luck!