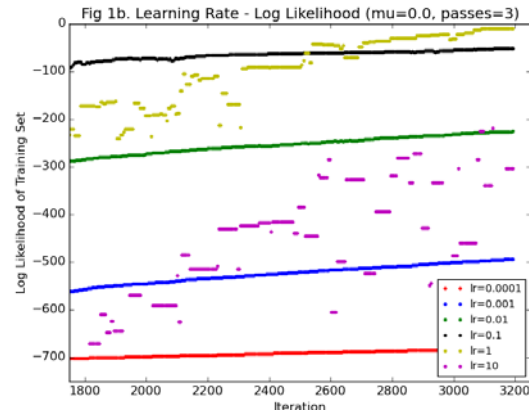
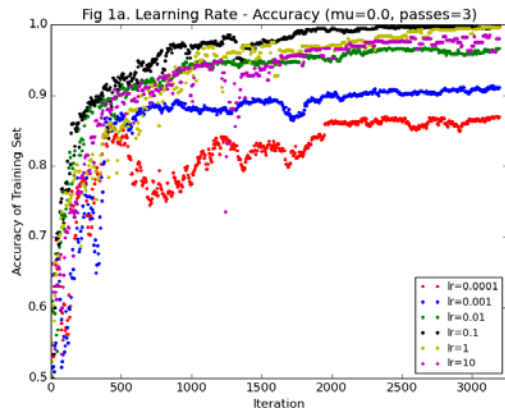
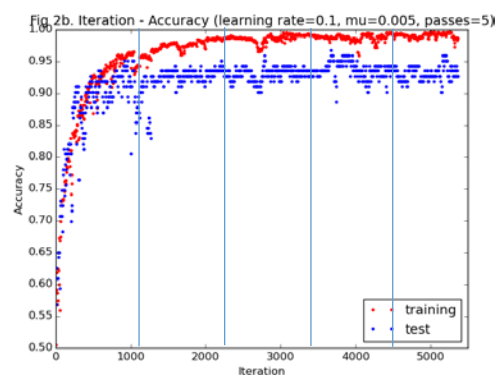
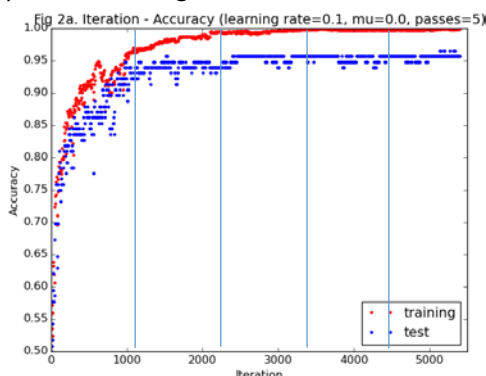


Homework 2 Analysis – Jianxiang Fan

- **Learning Rate** - Smaller learning rate means slower to converge, while too large rate tends to skip around and miss the optimal point. In my model, 0.1 is a pretty fair learning rate.



- **Passes** – The unregularized version usually needs 3 passes to get stable accuracy, while the regularized version needs more, 4 or 5 passes in average.



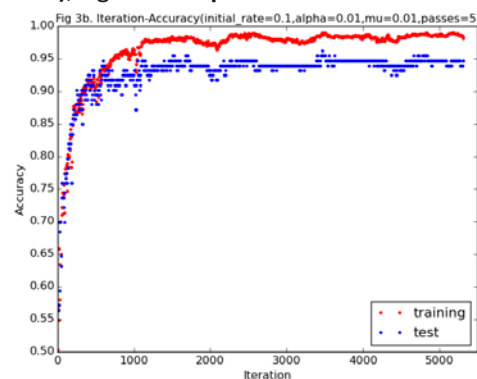
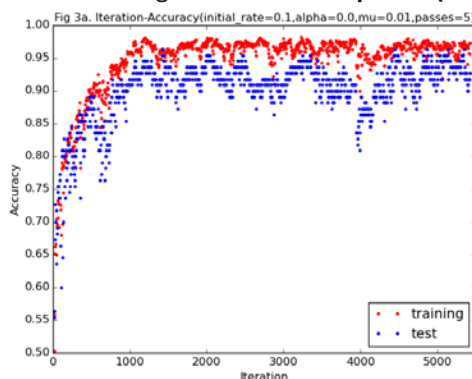
- **Best/Poorest Predictors** – Words with **largest positive** and **smallest negative** weights are the best predictors. And those words with **smallest absolute value (near 0)** weights are poorest predictors. Top 20 best and poorest:

Best for baseball: ['batting', 'happens', 'pitched', 'still', 'losing', 'steph', 'usa', 'stadium', 'dwarnar', 'bob', 'rickert', 'pitcher', 'luriem', 'ball', 'run', 'anyone', 'hit', 'baseball', 'pitching', 'runs']

Best for hockey: ['hockey', 'playoffs', 'next', 'points', 'playoff', 'golchowy', 'ice', 'goal', 'traded', 'pick', 'goals', 'finals', 'round', 'beat', 'gld', 'contact', 'biggest', 'need', 'coverage', 'predictions']

Poorest: ['blasted', 'hooked', 'silence', 'intermissions', 'deceased', 'tone', 'rode', 'riel', 'hesitate', 'racist', 'vintage', 'everywhere', 'memoriam', 'wrestling', 'pitiful', 'broad', 'bloody', 'retiring', 'privately']

- **Schedule to update the learning rate** – Update function: $\text{rate} = \text{initial_rate} / (1 + \alpha * \text{initial_rate} * \text{iteration})$. It helps to improve the convergence. Left one $\alpha=0$ (without update), right one $\alpha=0.01$



- **Use TF-IDF** – It makes training set more representative. Left one uses word frequency, right one uses tf-idf.

