Penn

▶ Define Statistical Risk Minimzation. Give examples of different losses

▶ Define Empiricial Risk Minimzation without parametrizations.

▶ Explain why this is a nonsensical formulation of machine learning.

▶ Define Empiricial Risk Minimzation with learning parametrizations.

▶ Exemplify with a linear parametriazation

▶ Exemplify with a neural network parametriazation

▶ What are the three components of an AI system?. Which is the only one that is a choice for the system designer? What is the property tat this choice controls?

▶ Define gradident descent and stochasstic gradient descent. Explain their differences

▶ Why does stochastic gradient descent convergence?

▶ Try to expand the squared norm $\|\mathbf{H}_{t+1} - \mathbf{H}^*\|^2 = \|\mathbf{H}_t - \epsilon\hat{\mathbf{g}}(\mathbf{H}_t) - \mathbf{H}^*\|^2$ and see if you find something interesting

▶ Explain the limit infimum convergence of stochastic gradient descent

▶ If the learning parametrization is matched to the underlying model we expect learning to work. Explain.

► If the learning parametrization is not matched to the underlying model we do not expect learning to work. Explain.

▶ If the learning parametrization is matched to the underlying model we expect learning to work only if we have sufficient data relative to the complexity of the problem. Explain.