Semi-Supervised Learning

Sargur N. Srihari
srihari@buffalo.edu

Regularization Strategies

1. Parameter Norm Penalties
2. Norm Penalties as Constrained Optimization
3. Regularization and Under-constrained Problems
4. Data Set Augmentation
5. Noise Robustness
6. Semi-supervised learning
7. Multi-task learning
8. Early Stopping
9. Parameter tying and parameter sharing
10. Sparse representations
11. Bagging and other ensemble methods
12. Dropout
13. Adversarial training
14. Tangent methods

Task of Semi-supervised Learning

- Both unlabeled examples from $P(\boldsymbol{x})$ and labeled examples from $P(\boldsymbol{x},y)$ are used to estimate $P(y|\boldsymbol{x})$ or predict y from $\boldsymbol{x}$
- In the context of deep learning it refers to learning a representation $\boldsymbol{h} = f(\boldsymbol{x})$
- The goal is to learn a representation so that examples from the same class have similar representations

How unsupervised learning helps

- Unsupervised learning can provide useful clues for how to group examples in representational space
- Examples that cluster tightly in the input space should be mapped to similar representations
- A linear classifier in the new space may achieve better generalization
- A variant is the application of PCA as a preprocessing step before applying a classifier to the projected data

Sharing Parameters

- Instead of separate unsupervised and supervised components in the model, construct models in which generative models of either $P(\boldsymbol{x})$ or $P(\boldsymbol{x}, y)$ shares parameters with a discriminative model of $P(y|\boldsymbol{x})$
- One can then trade-off the supervised criterion $-\log P(y|\boldsymbol{x})$ with the unsupervised or generative one (such as $-\log P(\boldsymbol{x})$ or $-\log P(\boldsymbol{x}, y)$)
  - The generative criterion then expresses a prior belief about the solution to the supervised problem
    - viz., structure of $P(\boldsymbol{x})$ is connected to structure of $P(y|\boldsymbol{x})$ in a way that is captured by shared parameterization