ECE595 / STAT598: Machine Learning I
Lecture 31.3: Regularization - Choosing a Regularization

Spring 2020
Stanley Chan

School of Electrical and Computer Engineering
Purdue University
Today's Lecture:

- Motivation for Regularization
  - VC Analysis
  - Example
- Two Regularization Techniques
  - Weight Decay
  - Augmented Error
- Choosing a Regularization
  - Pill or Poisson?
  - Role of $\lambda$
Choosing a Regularization: Pill or Poisson?

- Regularization = choose $\Omega(h)$ and $\lambda$.
- Choice of $\Omega(h)$ is heuristic.
- Finding a perfect $\Omega$ is as difficult as finding a perfect $\mathcal{H}$.
- Some forms of regularization work and some do not.
- Why bother with regularization if so many choices can go wrong?
- Regularization is a necessary evil.
- If our model is too sophisticated for the amount of data we have, we are doomed.
- By applying regularization, we have a chance.
Consider a 15-th order polynomial. So $\mathcal{H}_{15}$.

Two choices of regularization:
- Uniform regularization: $\Omega_{\text{uniform}}(\mathbf{w}) = \sum_{q=0}^{15} w_q^2$
- Low-order regularization: $\Omega_{\text{low}}(\mathbf{w}) = \sum_{q=0}^{15} q w_q^2$

When $\lambda$ too small, overfit. When $\lambda$ too large, underfit.

For optimal $\lambda$, the two are quite similar.
Regularization on Noise and Target Complexity

Let us analyze the impact of regularization to noise:

- **Noise**: Uncertainty in each measured data. Measured in terms of $\sigma^2$.
- **If** you have noise, **then** you need to adjust $\lambda$ depending on the noise level.
- **Target complexity**: Suppose data comes from $\mathcal{H}_{15}$ but you use $\mathcal{H}_{50}$. Measured in terms of $Q_f$.
- **Like** noise, **you need to adjust $\lambda$ to optimize generalization.**
What if Picked a Wrong Regularization?

- Suppose we should encourage low-order coefficients, but the regularization promotes high-order coefficients.
- Are we screwed?
- No, you still have the regularization parameter $\lambda$.
- Below is an example.
- Choosing the regularization parameter can be done using validation. Will discuss next.
Summary

- Whenever you train a model, try including regularization.
- It can be as simple as $w^T w$.
- Helps dramatically when there is noise in data, not enough data, complex target.
- Hand-waving argument: noise is high frequency. Complex target is also high frequency.
- So low-frequency regularization helps.
- As long as you have a good $\lambda$, the benefit of regularization is often more than the harm.
- Modern deep learning can easily incorporate regularization.
- E.g., you can regularize the magnitude of the network weights, or number of non-zeros through sparsity.
Reading List

- Yaser Abu-Mostafa, Learning from Data, chapter 4.2