Overfitting and Remedies

Find the perfect model complexity, Early stopping, Regularization, Dropout, Data augmentation, and Domain randomization
Outline

• Drawing recap for initialization and normalization

• Overfitting and its causes

• Overfitting remedies
  • Find the perfect model complexity
  • Early stopping
  • Regularization
  • Dropout
  • Data augmentation
  • Domain randomization
Recap: Parameter and Gradient Values

• Take five minutes to draw
• Example: activations with and without proper initialization and normalization
Classroom Etiquette

• We all want to look *effortlessly smart* in front of our peers.
  
  • It’s a fool’s errand. I’ve noticed it a bit in the class. Might be due to class makeup

• I’ve built my teaching philosophy around the “gift of failure”
  
  • You need to give me wrong answers

  • You need to be unafraid of being wrong

  • You need to be ready to fail
Overfitting

When your model **learns/memorizes** the training data and not **some property** that is useful for inference. ("I’ve seen this input before... the answer is X.")
Causes of Overfitting

When your model learns/memorizes the training data and not some property that is useful for inference. ("I’ve seen this input before… the answer is X.")

• The model is too complex
  • Too many parameters
  • Too deep
  • Too wide
  • Too much memory
• Parameters are too large (large parameters lead to steep curves)
• The model was trained for too long
• The dataset was too small
We could theoretically find the perfect model complexity for each problem.
Remedy: Early Stopping and Checkpointing

We can use the learned parameters from before we detected overfitting
Remedy: Regularization

We can artificially constrain the parameter magnitudes
**Remedy: Dropout**

We can train the model in such a way that breaks memorization

```
model.train() vs model.eval()
```
Remedy: Data Augmentation

https://albumentations.ai/
“Illustration of our approach. An object detector is trained on hundreds of thousands of low-fidelity rendered images with random camera positions, lighting conditions, object positions, and non-realistic textures. At test time, the same detector is used in the real world with no additional training.”

— Tobin et al.
Summary

- Models can accidentally memorize the input data instead of learning some useful, general property

- We can prevent overfitting/memorization with several remedies

- Most remedies try to
  - Artificially limit the magnitude of parameter values (early stopping, regularization)
  - Add noise and randomness to the training process (dropout, augmentation, domain randomization)