Debugging ML
The (ideal) plots

- **Left Plot**: Loss vs. Iterations during training
  - Blue line: Train
  - Red line: Val
- **Right Plot**: Accuracy vs. Iterations during training
  - Blue line: Train
  - Red line: Val
Overfitting (mild)

- Validation accuracy does not improve as training loss goes down
Overfitting (bad)

• Validation accuracy *decreases* as training loss goes down
Underfitting
Underfitting vs overfitting

Underfitting

Overfitting

Model capacity

Loss

Train

Val

Underfitting

Overfitting
Optimization issues
Optimization issues: Large step size

Small step sizes cause slow learning, Large step sizes cause divergence
SGD with momentum

- Stochastic gradient is stochastic
- Can reduce variance using momentum
- Standard update:
  \[ \theta_{t+1} = \theta_t - \lambda \nabla \theta F(\theta_t) \]
- With momentum:
  \[ \theta_{t+1} = \theta_t - \lambda p_{t+1} \]
  \[ p_{t+1} = \gamma p_t + \nabla \theta F(\theta_t) \]
Optimization issues: convergence

![Graph showing loss and accuracy over iterations during training.](image)

- **Loss**
  - Blue line: Train
  - Red line: Val

- **Accuracy**
  - Blue line: Train
  - Red line: Val