Early Stop,Dropout

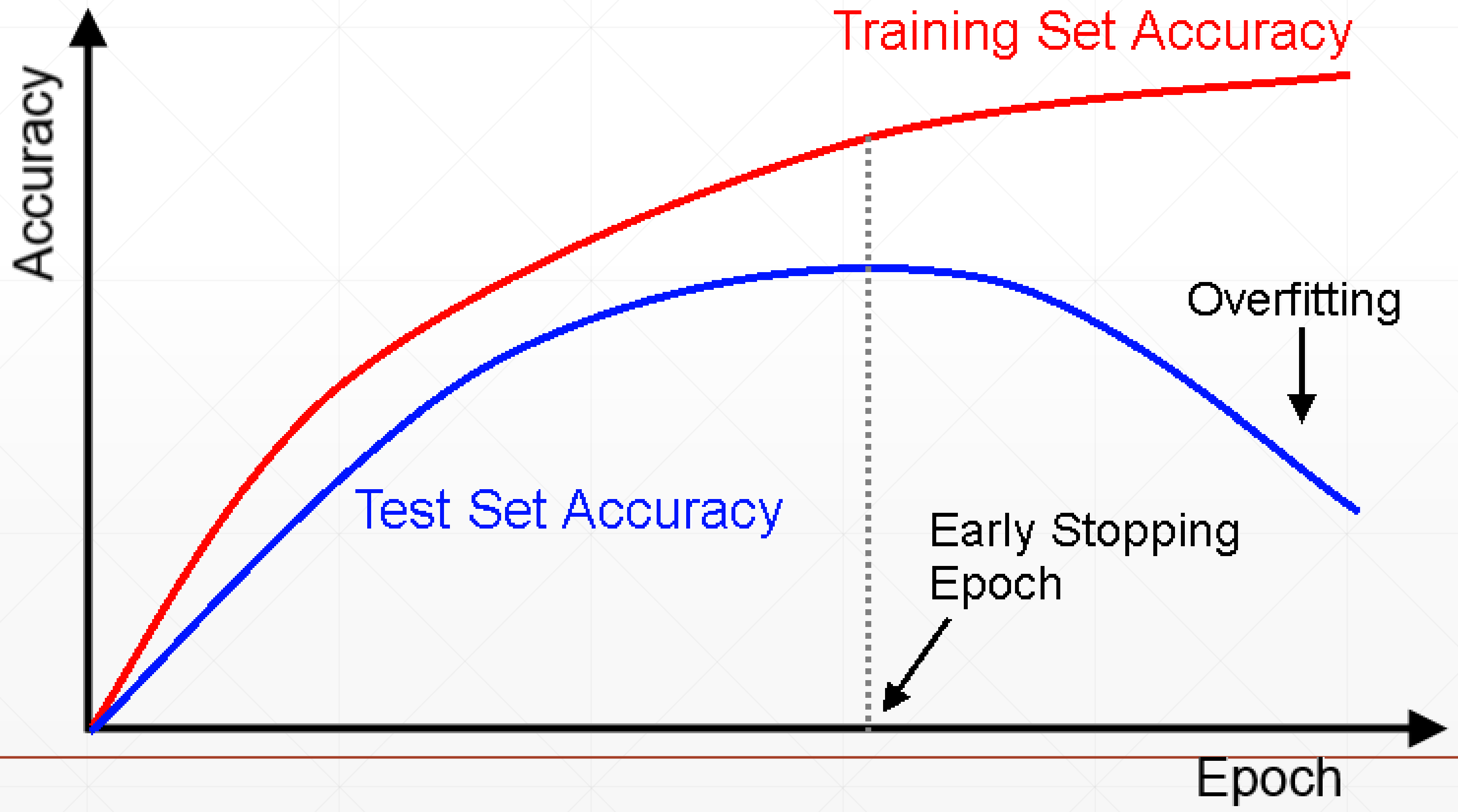
主讲人：龙良曲

Tricks

- Early Stopping
- Dropout
- Stochastic Gradient Descent

Early Stopping

- Regularization

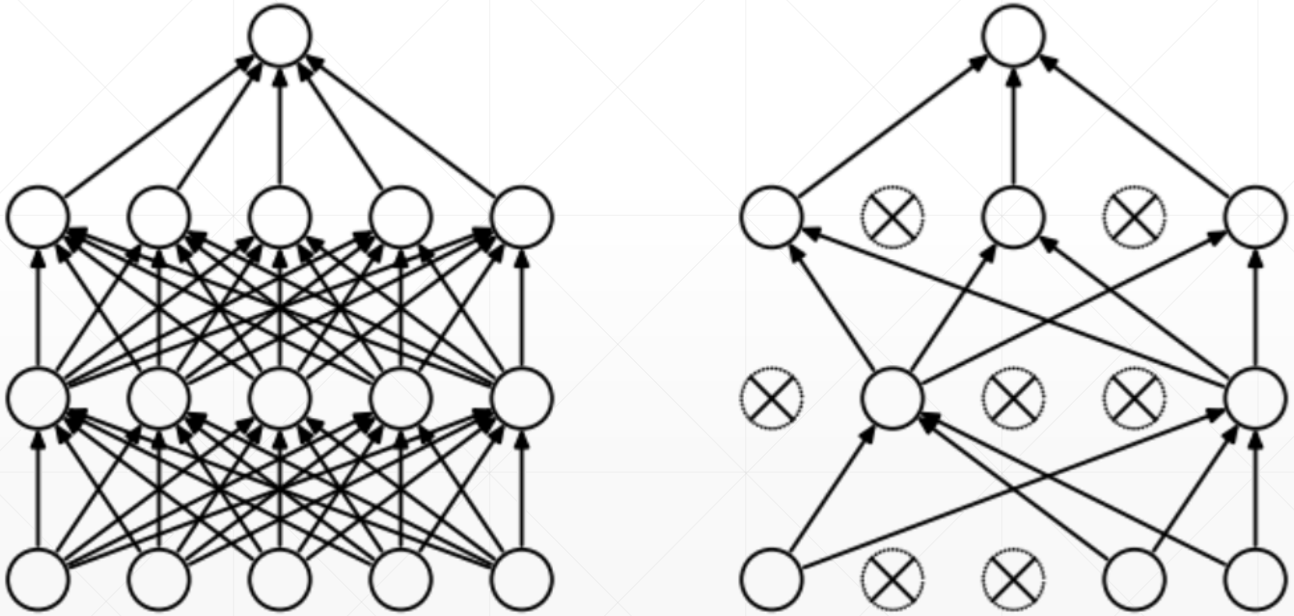


How-To

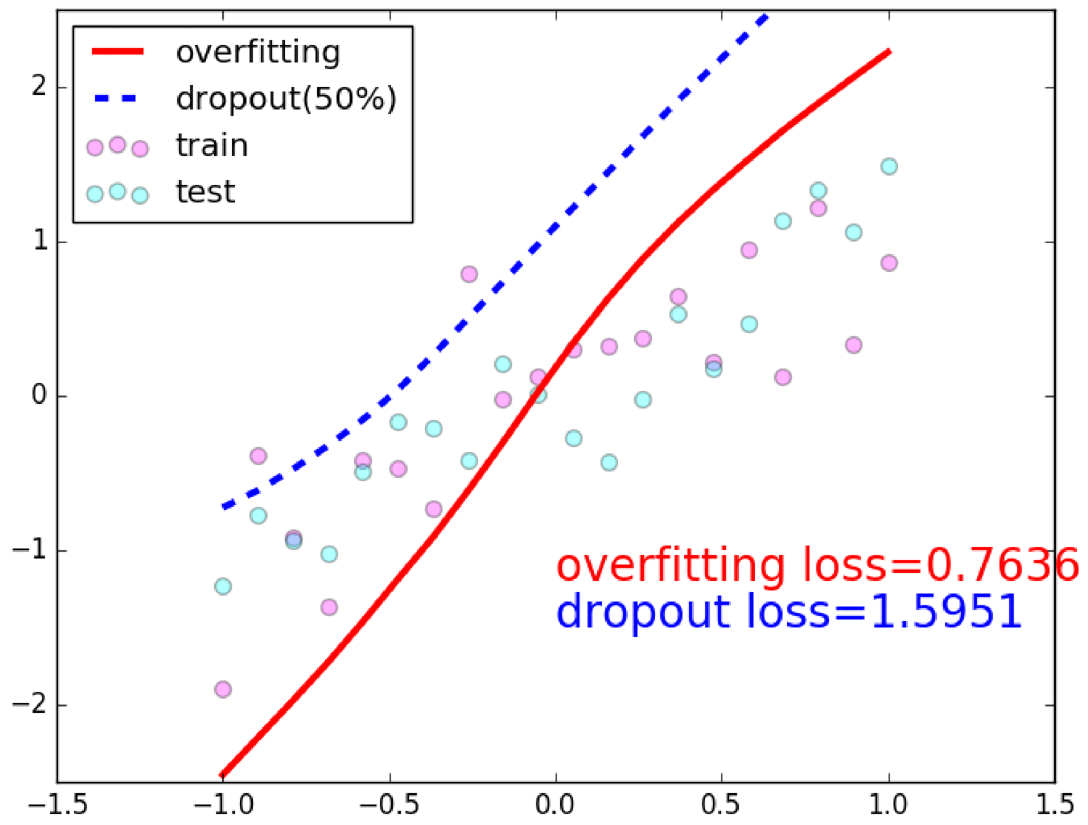
- Validation set to select parameters
- Monitor validation performance
- Stop at the highest val perf.

Dropout

- *Learning less to learn better*
- Each connection has $p = [0, 1]$ to lose



(a) Standard Neural Net (b) After applying dropout.



https://github.com/MorvanZhou/PyTorch-Tutorial

```python
net_dropped = torch.nn.Sequential(
    torch.nn.Linear(784, 200),
    torch.nn.Dropout(0.5),  # drop 50% of the neuron
    torch.nn.ReLU(),
    torch.nn.Linear(200, 200),
    torch.nn.Dropout(0.5),  # drop 50% of the neuron
    torch.nn.ReLU(),
    torch.nn.Linear(200, 10),
)
```

Clarification

- *torch.nn.Dropout(p=dropout_prob)*

- *tf.nn.dropout(keep_prob)*

Behavior between train and test

Batch-Norm

```
for epoch in range(epochs):

    # train
    net_dropped.train()
    for batch_idx, (data, target) in enumerate(train_loader):
      ...
    net_dropped.eval()
    test_loss = 0
    correct = 0
    for data, target in test_loader:
      ...
```

Stochastic Gradient Descent

- Stochastic
  - not random!
- Deterministic

Gradient Descent

$$\frac{\partial}{\partial \theta_j} J(\theta) = \frac{1}{m} \sum_{i=1}^{m} (\hat{y}^i - y^i) \cdot x_j^i$$

② Vanilla (Batch) G.D.

$$\theta_j := \theta_j - \alpha \cdot \underbrace{\frac{\partial}{\partial \theta_j} J(\theta)}_{\frac{1}{m} \sum_{i=1}^{m} (\hat{y}^i - y^i) x_j^i}$$

https://towardsdatascience.com/difference-between-batch-gradient-descent-and-stochastic-gradient-descent-1187f1291aa1

Gradient Descent

③ Stochastic G.D.

for i in range(m):

$$\theta_j := \theta_j - \alpha \cdot \boxed{\text{only one example}} \; \frac{\nabla J}{\nabla \theta_j}$$

$$(\hat{y}^i - y^i)\, x_j^i$$

https://towardsdatascience.com/difference-between-batch-gradient-descent-and-stochastic-gradient-descent-1187f1291aa1

Stochastic Gradient Descent

- Not single usually
- batch = 16, 32, 64, 128…

Why

¥1789.00

索泰（ZOTAC）GeForce GTX1060 X-GAMING OC吃鸡显卡/游戏电竞台式机独

6.6万+条评价 二手有售

索泰京东自营旗舰店

自营 秒杀 券1500-200

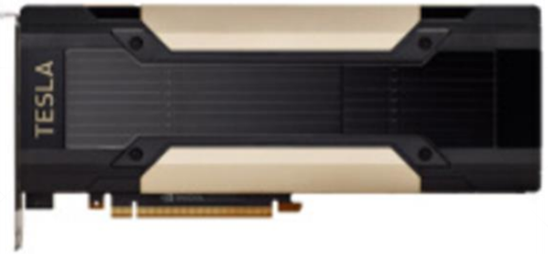
¥7599.00

索泰（ZOTAC） GTX1080Ti-11GD5X至尊PLUS OC绝地求生/吃鸡显卡

40+条评价

云轩 DIY整机专营店

满减 免邮 满199-100 险



¥80999.00

NVIDIA TESLA V100 16GB/32GB HBM2/人工智能深度学习GPU运算 16GB HBM2

0条评价

子希DIY电脑专营店

免邮

$$P(A|B) = \frac{P(B|A) \times P(A)}{P(B)},$$

下一课时

贝叶斯定理

Thank You.