

Embedding

embedding是一个分布式表征，最开始是由词向量被大家熟知，就是每一个词用向量来表示（如one_hot），但是这种方法在某些情况下（如高维），造成向量比较稀疏

如何理解embedding？其实可以理解为不带偏置的全连接神经网络，输入是one_hot向量，权重就是embedding table（单词数目 * embedding vector的深度）

如何得到embedding？把one_hot向量和embedding table相乘就可以得到每个单词的embedding vector

```
n, d, m = 3, 5, 7
embedding = nn.Embedding(n, d, max_norm=True) # 得到embedding table
w = torch.randn((m, d), requires_grad=True)
idx = torch.tensor([1, 2])
a = embedding.weight.clone() @ w.t() # weight must be cloned for this to be
differentiable
b = embedding(idx) @ w.t() # 把idx传入到embedding table中（这一步可以理解为全连接网络，
输入是index, 权重就是embedding table）
out = (a.unsqueeze(0) + b.unsqueeze(1))
loss = out.sigmoid().prod()
loss.backward()
```