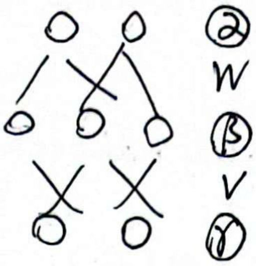
Chapter 5.

5.1

input layer

hidden-layer

output layer



如果激活函数用 $f(x) = W^T x$,

那么 $\beta = W^T \alpha$, $\gamma = V^T W^T \alpha$, 本质上是 linear regression

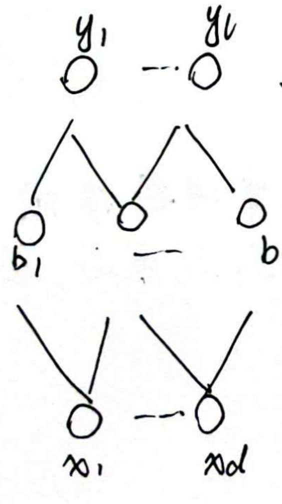
多 hidden layer 都是线性的,

换成 sigmoid 函数, 就变成 $\gamma = f(V^T f(W^T \alpha))$,

特征更多样化, 结构更复杂了.

5.2. ~~single~~ 无 hidden layer 的 FNN, 等价于 multi-class logistic regression.

5.3



output $\beta_j = \sum_{h=1}^{q} w_{hj} \hat{b}_h$, $\hat{y}_j = f(\beta_j)$

hidden $\alpha_h = \sum_{i=1}^{d} v_{ih} \cdot x_i$, $\hat{b}_h = f(\alpha_h)$

$\Delta v_{ih} = -\eta \frac{\partial E_k}{\partial v_{ih}}$, $\frac{\partial E}{\partial v_{ih}} = \sum_{j=1}^{l} \left(\frac{\partial E}{\partial \hat{y}_j} \cdot \frac{\partial \hat{y}_j}{\partial \beta_j} \right) \frac{\partial \beta_j}{\partial \hat{b}_h} \frac{\partial \hat{b}_h}{\partial \alpha_h} \cdot \frac{\partial \alpha_h}{v_{ih}}$

$$\frac{\partial E}{\partial \hat{y}_j} \cdot \frac{\partial \hat{y}_j}{\partial \beta_j} \cdot \frac{\partial \beta_j}{\partial b_h} = (\hat{y}_j - y_j) \cdot f(\beta_j)(1 - f(\beta_j)) \cdot w_{hj}$$

$$= (\hat{y}_j - y_j) \cdot \hat{y}_j (1 - \hat{y}_j) \cdot w_{hj} = -g_j \cdot w_{hj}.$$

$$\frac{\partial b_h}{\partial \alpha_h}\cdot\frac{\partial \alpha_h}{\partial v_{ih}} = (1-b_h)\cdot b_h\cdot x_i$$

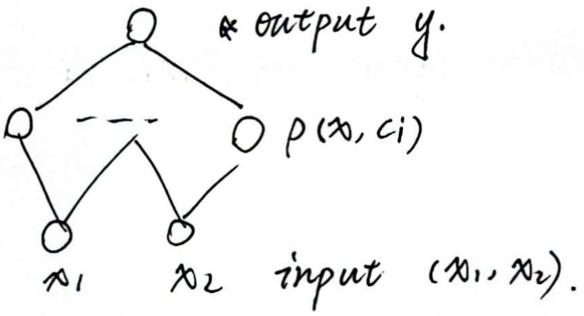
$$\Delta v_{ih} = +\eta\cdot\sum_{j=1}^{l} g_j\cdot w_{hj}\cdot b_h(1-b_h)\cdot x_i.$$

$$g_j = -\hat{y}_j(1-\hat{y}_j)(-y_j+\hat{y}_j). \quad \square$$

5.4. η learning rate 对收敛速度和收敛精度有影响.

5.7. 使用数据 ~~$(x, y) = 1$~~ $(x_1, x_2, y) = (0,0,0), (0,1,1), (1,0,1), (1,1,0)$

训练. Graph ↓



~~5.9~~ 5.9. 也一样用链式法则求导，不过中间要移项.
就是隐函数求导，差不多.