1. The sum of weights.

$$= 10 * d^{(1)} + (d^{(1)}+1)\, d^{(2)} + \cdots\cdots + (d^{(L-2)}+1)(d^{(L-1)}) + d^{(L-1)} \times 1$$

$$\text{s.t.} \sum_{t=1}^{L-1} d^{(t)}+1 = 36.$$

for minimum weight number

if $d^{(1)} = d^{(2)} = \cdots d^{(18)} = 1$

$$\therefore \text{min \# of weight} = \overbrace{10 + 2 + \cdots + 2 + 2}^{18} = 46$$

2. For maximum sum of weight numbers.

$\Rightarrow$ See program. greedy search.

Max = 510

3.

$$\nabla \text{err}_n(w) = \frac{\partial\, (x_n - ww^T x_n)^T (x_n - ww^T x_n)}{\partial w}$$

$$= \boxed{2\,(x_n - ww^T x_n)} \cdot \boxed{\frac{\partial\,(x_n - ww^T x_n)}{\partial w}}$$

$$= \cancel{-2\,(x_n - ww^T x_n)\,(x_n + x_n^T)\, w}$$

$$= -2\,(x_n + x_n^T)\, w\,(x_n - ww^T x_n)$$

$$= 2(x_n^T w)^2 w + 2(x_n^T w)(w^T w)\, x_n - 4 w^T x_n x_n$$

Double A

4. $\tilde{x}_n = x_n + \varepsilon_n$ $\quad E(\varepsilon_n) = 0_n$

$\varepsilon_n \sim N(0_n, I)$

$$E_{in} = \frac{1}{N}\sum_{n=1}^{N} (x_n - ww^T x_n)^T (x_n - ww^T x_n)^T$$

$$2\frac{1}{N}\sum (x_n - ww^T x_n)^T (ww^T \varepsilon_n)$$

$$+ \sum (ww^T \varepsilon_n)^T (ww^T \varepsilon_n)$$

$$\Omega(w) = E\Big[2\frac{1}{N}\sum (x_n - \cancelto{0}{w^T w^T x_n})^T (ww^T \varepsilon_n)$$
$$+ \frac{1}{N}\sum (ww^T \varepsilon_n)^T (ww^T \varepsilon_n) \Big]$$

$$= \frac{1}{N}\sum E[\varepsilon_n^T ww^T ww^T \varepsilon_n]$$

$$= w^T w \frac{1}{N}\sum E[\varepsilon_n^T ww^T \varepsilon_n)$$

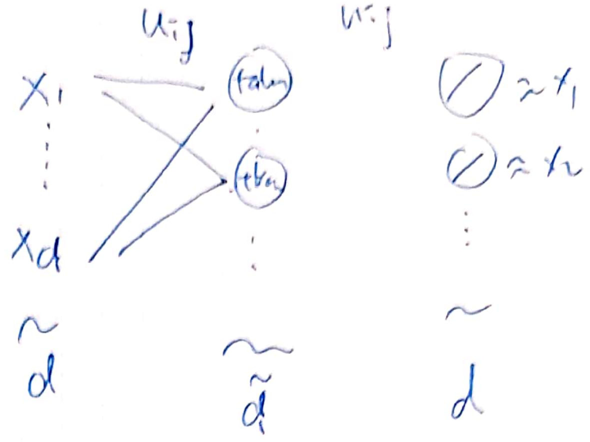
$$= w^T w \frac{1}{N}\sum E[\sum_{i=1}^{d} \varepsilon_{ni}^2 w_i^2]$$

$$= w^T w \frac{1}{N}\sum^{N}\sum^{d} w_i^2 E(\varepsilon_{ni}^2)$$

$$= w^T w \frac{1}{N}\sum^{N}\sum^{d} w_i^2 = \frac{1}{N}\sum (w^T w)^2 = \underline{(w^T w)^2}.$$

Double A

5.



error function

$$= \sum_{i=1}^{d} \left(g_i(x) - x_i \right)^2$$

$$= \sum_{j=1}^{d} \left(\sum_{k=1}^{\tilde{d}} \underbrace{u_{ki}}_{w_{ji}^{(2)}} \tanh\left(\sum_{i=1}^{d} \underbrace{u_{ij}}_{w_{ij}^{(1)}} x_i \right) - x_j \right)^2$$

6.

$$\frac{\partial E}{\partial u_{ij}} = 2 \sum^{d} \left(\sum^{\tilde{d}} u_{ki} \tanh\left(\sum^{d} u_{ij} x_i \right) - x_j \right) \left[\tanh\left(\sum^{d} u_{ij} x_i \right) + \sum_{i=1}^{\tilde{d}} u_{ki} \tanh\left(\sum u_{ij} x_i \right) \tanh'\left(\sum u_{ij} x_i \right) x_i \right]$$

$$\frac{\partial E}{\partial w_{ij}^{(1)}} = 2 \sum^{d} \left(\sum^{d} u_{ki} \tanh\left(\sum u_{ij} x_i \right) - x_j \right) \cdot \left(\sum^{\tilde{d}} u_{ki} \tanh\left(\sum u_{ij} x_i \right) \right) x_i$$

$\tanh'\left(\sum u_{ij} x_i \right)$

$+$

$$\frac{\partial E}{\partial w_{ij}^{(2)}} = 2 \sum^{d} \left(\sum^{d} u_{ki} \tanh\left(\sum u_{ij} x_i \right) - x_j \right) \cdot \tanh\left(\sum w_{ij}^{(1)} x_i \right)$$

$\parallel$

$$2 \sum^{d} \left(\sum^{d} u_{ki} \tanh\left(\sum u_{ij} x_i \right) - x_j \right) \left[\sum^{\tilde{d}} u_{ki} \tanh\left(\sum u_{ij} x_i \right) \tanh'\left(\sum u_{ij} x_i \right) x_i + \tanh\left(\sum w_{ij}^{(1)} x_i \right) \right]$$

7.

$$g_{LIN}(x) = \text{sign}\left(\|X - X_-\|^2 - \|X - X_+\|^2 \right)$$

$$= \text{sign}\left((X - X_-)^T (X - X_-) - (X - X_+)^T (X - X_+) \right)$$

$$= \text{sign}\left(2(X_+ - X_-)^T X + \|X_-\|^2 - \|X_+\|^2 \right)$$

8. if $g_{RBFNET}(x) \geq 1$

$$\Rightarrow \beta_+ \exp(-\|X - \mu_+\|^2) + \beta_- \exp(-\|X - \mu_-\|^2) \geq 0$$

$$\Rightarrow \frac{\exp(-\|X - \mu_+\|^2)}{\exp(-\|X - \mu_-\|^2)} \geq -\frac{\beta_-}{\beta_+}$$

$$\Rightarrow \underline{\|X - \mu_-\|^2 - \|X - \mu_+\|^2} \geq \ln\left(-\frac{\beta_-}{\beta_+}\right)$$

as g_{LIN}, using (7) solution

$$W = 2(\mu_+ - \mu_-)$$

$$b = \|\mu_-\|^2 - \|\mu_+\|^2 - \ln\left(-\frac{\beta_-}{\beta_+}\right)$$

9.

$$\sum_{n=1}^{N} \sum_{(x_n, r_{nm}) \in D_m} (r_{nm} - w_m^T v_n)^2$$

FOC:

$$2 \sum_{(x_n, r_{nm}) \in D_m} (r_{nm} - w_m^T v_n) v_n = 0$$

$$\Rightarrow \sum (r_{nm} - w_m^T v_n) = 0$$

$$\Rightarrow w_m = \frac{\sum_n r_{nm}}{\sum_n v_n} = \underbrace{\frac{\sum_n r_{nm}}{n}}_{\text{average rating}}.$$

10.

$$\max_m v_{n+1}^T v_m = \max_m \left(\frac{1}{N} \sum v_n\right)^T w_m$$

$$= \max_m \frac{1}{N} \sum v_n^T w_m = \max \underbrace{\frac{1}{N} \sum^{N} r_{nm}}_{\text{maximum of average rating}}.$$