② 
$$MSE = \frac{1}{N}\left[\sum_{x=c_1}[f(x,\theta)-a]^2 + \sum_{x=c_2}[f(x,\theta)-b]^2\right]$$

let $a = 1$ and $b = 0$

$$MSE = \frac{1}{N}\left[\sum_{x=c_1}[f(x,\theta)-1]^2 + \sum_{x=c_2}[f(x,\theta)\cdot]^2\right]$$

for M > 2 classes, sum the outputs samples
and divide by ~~expects~~

$$MSE = \frac{1}{N}\sum_{k=1}^{M}\left[\sum_{x_i\in\omega_k}[f_k(x_i,\theta)-1]^2 + \sum_{x_i\notin\omega_k}f_k^2(x_i,\theta)\right]$$

we use $x_i$ for any input feature,

this is the MSE for any network that
has more than ~~the~~ two classifiers
given that the output of the correct
class is 1 and 0 for all the other
classifiers.