1. We start with the fours basis vectors:
$r^1 = [-1,-1]^T$
$r^2 = [1,1]^T$
$r^3 = [-1,1]^T$
$r^4 = [1,-1]^T$

We then transform every two dimensional vector x to a four dimensional vector such that:
$f(x) = \exp(-(x - r_i)^2)$
We now see that the ith element of the four dimensional vector is inversely proportional to the distance between the starting two dimensional vector and the ith corresponding basis vector. Thus the output will be closer to 1 when the input and the corresponding basis vector are similar to each other and close in terms of distance, and 0 when they are not. This transformation solves the XOR problem by making it linearly separable through the weight matrix: $w = [1, 1, -1, -1, 0]^T = [y_1, y_2, y_3, y_4, 0]^T$ with y_i being the class label of the i^{th} basis vector. In this case r^1 and r^2 represent y = 1 while r^3 and r^4 represent the negative class y = -1. We can conclude that input vectors in the positive class will have their first or second element of their transformed vector close to 1 and input vectors of the negative class will have the third or fourth element of their transformed vector close to 1, which means that the input vector belongs to the class that its nearest basis vector belongs to and that the weight matrix solves the XOR problem.

2. We define K = number of classes in the multi-call classification problem. To create a weight matrix we first select a K basis vectors with the corresponding labels in the following form: $\{(r_1, y_1), \ldots (r_k, y_k)\}$ and transform the input vectors in a matrix containing the transformed vector where row I corresponds to input vector i such that: $f(x) = [\exp(-(x-r_i)^2]$ and derive the weight matrix from it with the same logic followed above such that $w = [y_1, \ldots, y_k, 0]^T$ with y_i containing the one-hot vector to which the i^{th} basis vector belongs.

3. $D(M(\phi(x)), M, \phi(x)) = -(y^*\log(M(\phi(x)))+ ((1-y^*)(\log(1-M(\phi(x))))$
we set $a = -y^*\log(M(x))$ and $b = -(1-y^*)\log(1-M(x))$
Plugging in for M(x) we get:
$a = -y^*\log(\sigma(w^T\tilde{x})) = -y^*\log(\frac{1}{1+e^{-w^T\tilde{x}}})$

$b = -(1-y^*)\log(1-\sigma(w^T\tilde{x})) = (1-y^*)\log(1- \frac{1}{1+e^{-w^T\tilde{x}}}) = (1-y^*)\log(\frac{1+e^{-w^T\tilde{x}}}{1+e^{-w^T\tilde{x}}} - \frac{1}{1+e^{-w^T\tilde{x}}}) = (1-y^*)\log(\frac{e^{-w^T\tilde{x}}}{1+e^{-w^T\tilde{x}}}) = (1-y^*)(\log(e^{-w^T\tilde{x}}) - \log(1+e^{-w^T\tilde{x}}))$

$$\frac{\partial a}{\partial w^T} = -\frac{y\tilde{x}*e^{-w^T\tilde{x}}}{1+e^{-w^T\tilde{x}}}$$

$$\frac{\partial b}{\partial w^T} = -(1-y^*)\left(-\frac{\tilde{x}*e^{-w^T\tilde{x}}}{e^{-w^T\tilde{x}}} - -\frac{\tilde{x}*e^{-w^T\tilde{x}}}{1+e^{-w^T\tilde{x}}}\right) = (1-y*)\left(\frac{-\tilde{x}\left(1+e^{-w^T\tilde{x}}\right)+\tilde{x}e^{-w^T\tilde{x}}}{1+e^{-w^T\tilde{x}}}\right)$$
$$= -(1-y^*)(\frac{-\tilde{x}}{1+e^{-w^T\tilde{x}}})$$

$\nabla rkD\left(y^{*}, M, \phi(x)\right) = -\frac{y\tilde{x}*e^{-w^{T}\tilde{x}}}{1+e^{-w^{T}\tilde{x}}} - (1-y^{*})\left(\frac{-\tilde{x}}{1+e^{-w^{T}\tilde{x}}}\right) = \frac{-y^{*}\tilde{x}e^{-w^{T}\tilde{x}}+\tilde{x}-y^{*}\tilde{x}}{1+e^{-w^{T}\tilde{x}}} = \frac{\tilde{x}(-y^{*}e^{-w^{T}\tilde{x}}+1-y^{*})}{1+e^{-w^{T}\tilde{x}}}$

since $\sigma(w^{T}\tilde{x}) = \frac{1}{1+e^{-w^{T}\tilde{x}}}$

$= \sigma(w^{T}\tilde{x})\tilde{x}\left(-y^{*}e^{-w^{T}\tilde{x}} + 1 - y^{*}\right) = (-y^{*}\tilde{x} + \tilde{x}\sigma(W^{T}\tilde{x}) = \tilde{x}(-y^{*} + \sigma(w^{T}\tilde{x}))$

Setting $\phi(x) = x$

$\nabla rk\left(\phi(x)\right) = -2w\_k\phi\_k(x)(x - r^{k})$ and then substituting $\nabla rk\left(\phi(x)\right)\ for\ \tilde{x}$

$\nabla rkD\left(y^{*}, M, \phi(x)\right) = -2(y^{*} - \sigma\left(w^{T}\phi(x)\right))(w_{k}\phi_{k}(x)(x - r^{k}))$