2. $X = 3.2$

closest are $(3,5)$; $(2,11)$; $(3,8)$

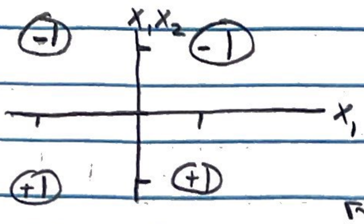
$$g(\vec{x}) = \frac{1}{k}\sum_{i=1}^{k} y_{[i]}(\vec{x})$$
$$= \frac{1}{3}(5+11+8) = \boxed{8}$$

3.

x_1	x_2	XOR
-1	-1	-1
+1	-1	+1
-1	+1	+1
+1	+1	-1

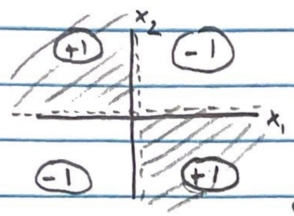
$\xrightarrow{\text{MAP}}$

x_1	x_1x_2	XOR
-1	+1	-1
+1	-1	+1
-1	-1	+1
+1	+1	-1



- Max margin is at $x_1x_2 = 0$.
- Margin is 1.

Transform space



- Seperator is at $x_1 = 0$, $x_2 = 0$.
- kind of like hyperbola.

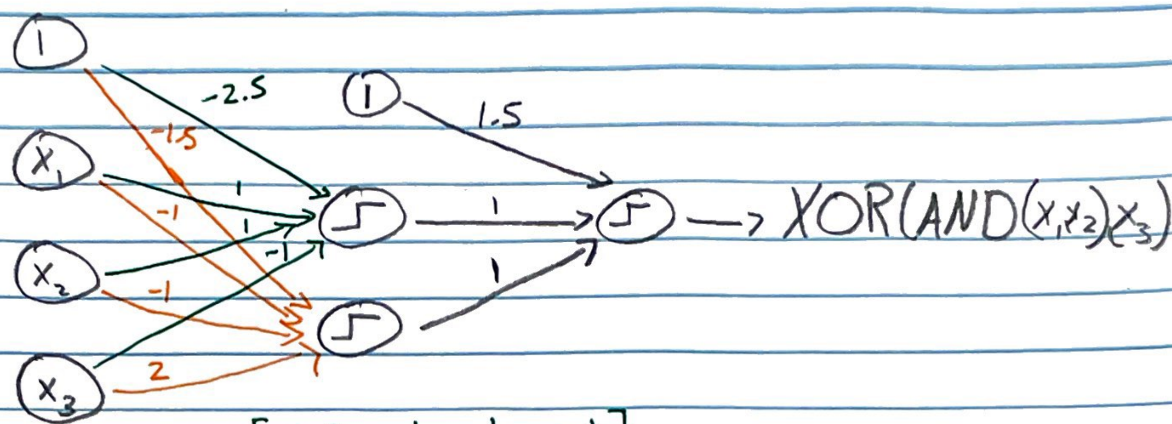
original space

4. $$\text{dist} = \|\phi(\vec{x}_i) - \phi(\vec{x}_j)\|^2$$
$$= \phi(\vec{x}_i)^2 - 2\phi(\vec{x}_i)^T\phi(\vec{x}_j) + \phi(\vec{x}_j)^2$$
$$= \boxed{K(\vec{x}_i, \vec{x}_i) - 2K(\vec{x}_i, \vec{x}_j) + K(\vec{x}_j, \vec{x}_j)}$$

5. $XOR(AND(x_1, x_2), x_3)$

make $AND(x_1, x_2)$ be x_{12}

$XOR(x_1 x_2, x_3) \rightarrow x_1 x_2 \bar{x}_3 + \overline{x_1 x_2}\, x_3$



$w_1 = [-2.5, \ 1, \ 1, \ -1]$

$w_2 = [-1.5, \ -1, \ -1, \ 2]$

$w_3 = [1.5, \ 1, \ 1]$

6a. $E_{in}(\vec{w}) = \frac{1}{N}\sum_{n=1}^{N}\left(\tanh(\vec{w}^T\vec{x}_n) - y_n\right)^2$

$\nabla E_{in}(\vec{w}) = \frac{1}{N}\sum_{n=1}^{N} 2\left(\tanh(\vec{w}^T\vec{x}_n) - y_n\right)\left(\frac{\partial}{\partial w}\left(\tanh(\vec{w}^T x_n) - y_n\right)\right)$

derivitive of $\tanh(x)$ is $1-\tanh^2(x)$

$\boxed{\nabla E_{in}(\vec{w}) = \frac{2}{N}\sum_{n=1}^{N}\left(\tanh(\vec{w}^T\vec{x}_n) - y_n\right)\left(1-\tanh^2(\vec{w}^T x_n)\right)\vec{x}_n}$

If $\vec{w}$ goes to ∞, then tanh will basically be 1 always, and there will be basically no gradient in any direction. This will make it difficult to improve weights.

6b. If all the weights are 0, then $\tanh(0)=0$ for the output of all layers. Then, all layers except the input are 0 and the $\vec{x}_n$ in the gradient update will set the gradient to 0! Thus there will never be a direction for us to update since the grad will always be 0.