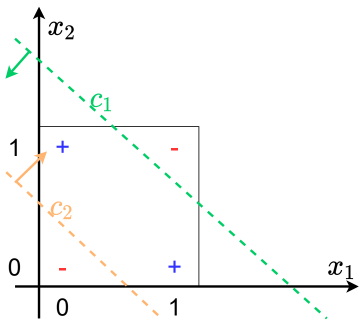
Q1

Given in Lecture Note 10.

Q2

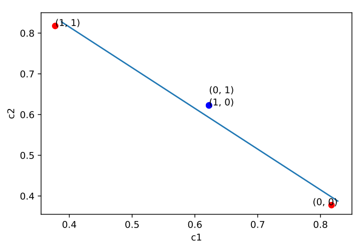
We have two linear classifiers c_1 and c_2 positioned as following,



	(0, 0)	(0, 1)	(1, 0)	(1, 1)
c_1	+	+	+	-
c_2	-	+	+	+

We can instantiate classifiers c_1 and c_2 as following, $c_1 = \sigma(-x_1 - x_2 + 1.5)$, $c_2 = \sigma(x1 + x2 - 0.5)$.

Project them onto the space of c_1 and c_2,



	(0, 0)	(0, 1)	(1, 0)	(1, 1)
c_1	0.8176	0.6225	0.6225	0.3778
c_2	0.3778	0.6225	0.6225	0.8176

We can have the third classifier as $y = \sigma(c_1 + c_2 - 1.22)$.

	(0, 0)	(0, 1)	(1, 0)	(1, 1)
y	0.4938	0.5062	5062	4938

Q3

In Q2, we have the following weights:

$$
\begin{aligned}
c_1 : & \boldsymbol{w}_1 = (-1, -1, 1.5) \\
c_2 : & \boldsymbol{w}_2 = (1, 1, -0.5) \\
y : & \boldsymbol{w}_3 = (1, 1, -1.22)
\end{aligned}
$$

Now let us swap the parameters of c_1 and c_2, and we have a new model as the following,

$$\begin{aligned} c_1' :& \boldsymbol{w}_1' = (1, 1, -0.5) \\ c_2' :& \boldsymbol{w}_2' = (-1, -1, 1.5) \\ y' :& \boldsymbol{w}_3' = (1, 1, -1.22) \end{aligned}$$

It is easy to verify that $y' = y$, which means y' is also a valid solution.

Average the weights $\tilde{w}_1 = \frac{\boldsymbol{w}_1 + \boldsymbol{w}_1'}{2}$, $\tilde{w}_2 = \frac{\boldsymbol{w}_2 + \boldsymbol{w}_2'}{2}$, and $\tilde{w}_3 = \frac{\boldsymbol{w}_3 + \boldsymbol{w}_3'}{2}$ and obtain another model

$$\begin{aligned} \tilde{c_1} :& \tilde{\boldsymbol{w}}_1 = (0, 0, 0.5) \\ \tilde{c_2} :& \tilde{\boldsymbol{w}}_2 = (0, 0, 0.5) \\ \tilde{y} :& \tilde{\boldsymbol{w}}_3 = (1, 1, -1.22) \end{aligned}$$

	(0, 0)	(0, 1)	(1, 0)	(1, 1)
$\tilde{y}$	0.5062	0.5062	5062	5062

Suppose the positive data points $(0, 1)$, $(1, 0)$ are labeled and 1, and the negative data points $(0, 0)$ and $(1, 1)$ are labeled as 0.

The loss for the y and y' model is $\mathcal{L} = -\log(0.4938) - \log(0.4938) - \log(1 - 0.5062) - \log(0.5062) \approx 2.723$. However, the loss for $\tilde{y}$ is $\tilde{\mathcal{L}} = -\log(0.5062) - \log(0.5062) - \log(1 - 0.5062) - \log(0.5062) \approx 2.773 > \mathcal{L}$.

If the optimization of the XOR model is convex, we should have the loss on the averaged weight smaller, i.e., $\mathcal{L} > \tilde{\mathcal{L}}$. This contradiction shows that the optimization is non-convex.