

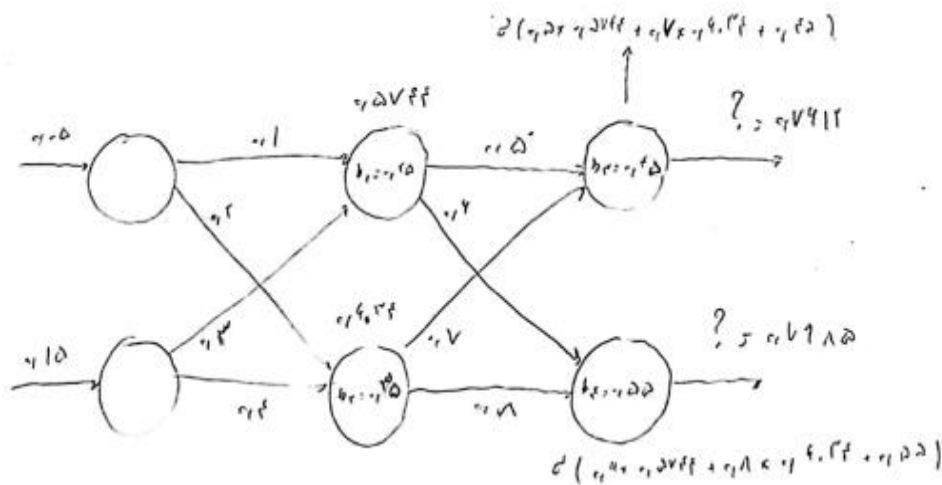
# **Introduction to Artificial Intelligence**

**Mahdi Naeeni - 810101536**

**HW5**

## بهینه سازی

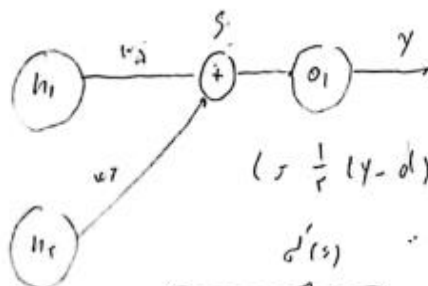
### سوال اول



(الف)

$$E = \frac{1}{2} \left[ (y_1 - v_{1112})^2 + (y_2 - v_{1118})^2 \right] = 0.125$$

$$w_{11}^{t+1} = w_{11}^t - \eta \frac{\partial L}{\partial w_{11}}$$



(ج)

$$\left( \frac{1}{2} (y - d)^2 \right) \Rightarrow \frac{\partial L}{\partial y} = y - d$$

$$\frac{\partial L}{\partial w_{11}} = \frac{\partial L}{\partial y} \cdot \frac{\partial y}{\partial s} \cdot \frac{\partial s}{\partial w_{11}} = (y - d) \sigma'(s) (1 - \sigma(s)) h_1$$

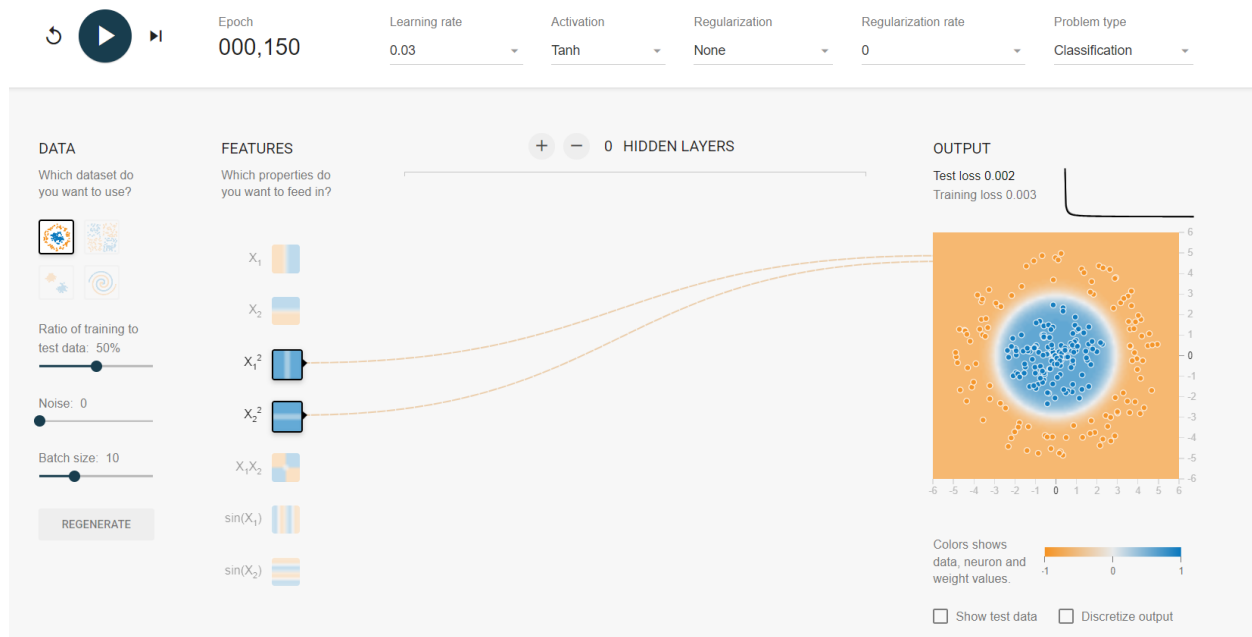
$$\frac{\partial L}{\partial w_{11}} = (0.1112 - 0.1) (0.1112) (0.1 - 0.1112) 0.125 = 0.00001$$

$$\Rightarrow w_{11}^{t+1} = w_{11}^t - \frac{1}{2} \times (0.1112) = w_{11}^t - 0.00001 = 0.125$$

## شبکه های عصبی

### سوال اول

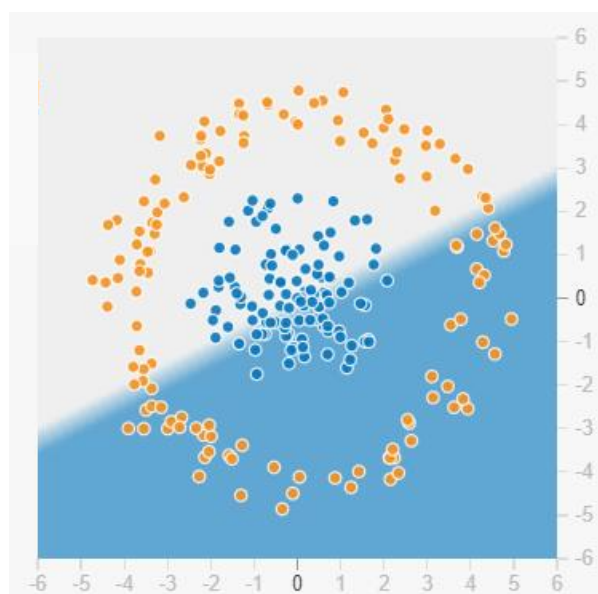
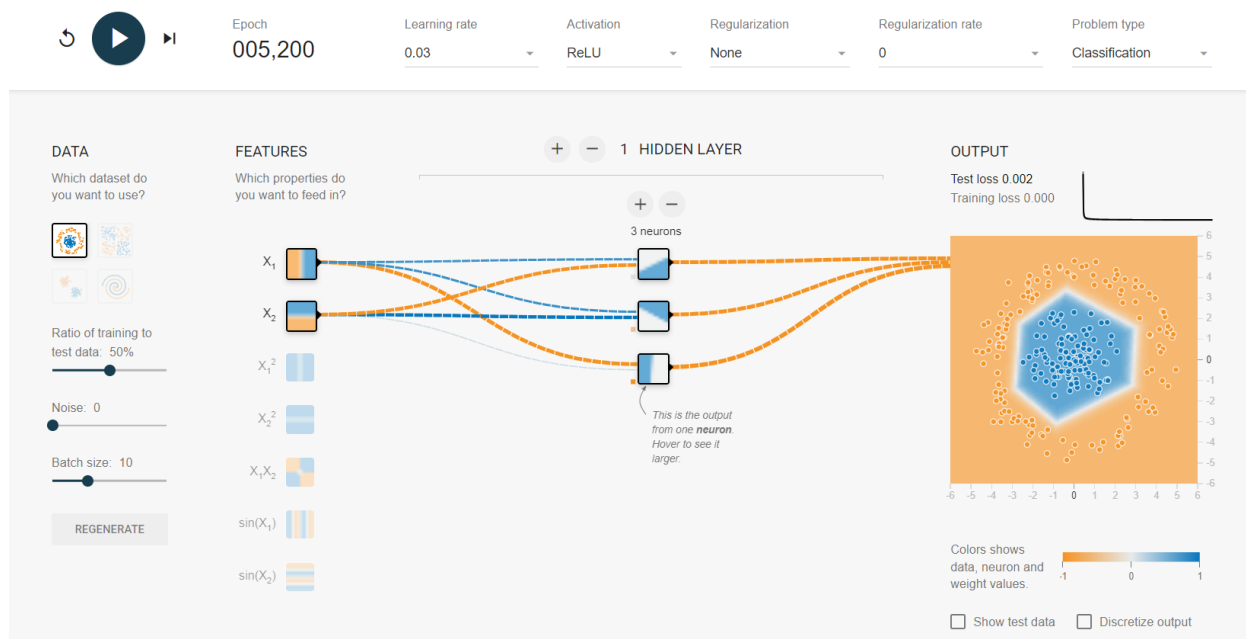
(الف)



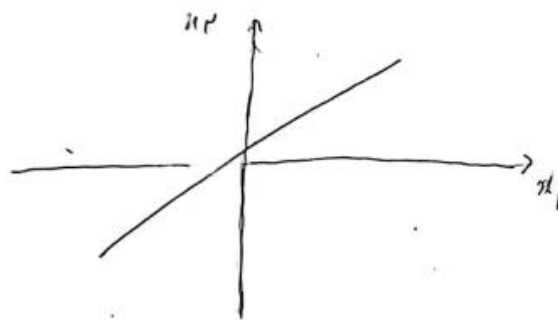
شبکه در این سناریو به درستی کار می کند زیرا ویژگی های  $X_1^2$  و  $X_2^2$  برای جمع آوری اطلاعات لازم برای طبقه بندی داده های دایره ای کافی هستند.

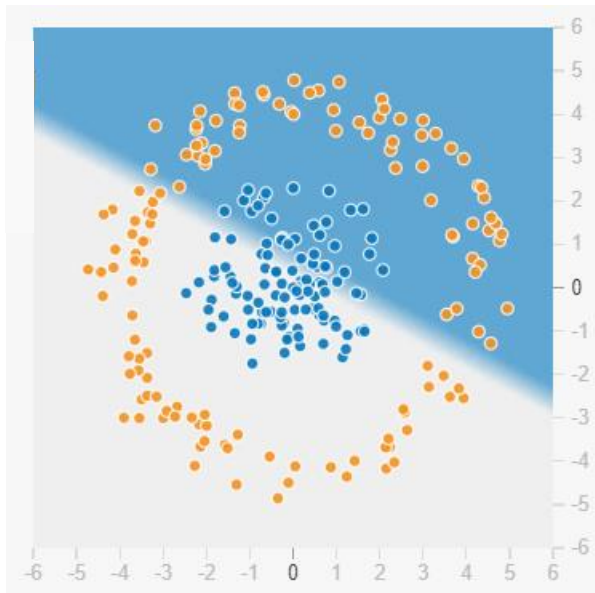
داده ها دایره ای هستند، به این معنی که مرز تصمیمی که دو کلاس را از هم جدا می کند، اساساً یک دایره است. با استفاده از  $X_1^2$  و  $X_2^2$  به عنوان ویژگی ها، مدل به راحتی می تواند مرز تصمیم گیری دایره ای را بدست آورد. مربع کردن مختصات  $X_1$  و  $X_2$  آنها را به فضایی تبدیل می کند که مرز تصمیم دایره ای به یک خطی تبدیل می شود. در این مورد، یک مدل خطی (بدون لایه های پنهان) می تواند به طور موثر داده ها را جدا کند.

$$w_1 * x_1^2 + w_2 * x_2^2 = w_0^2$$

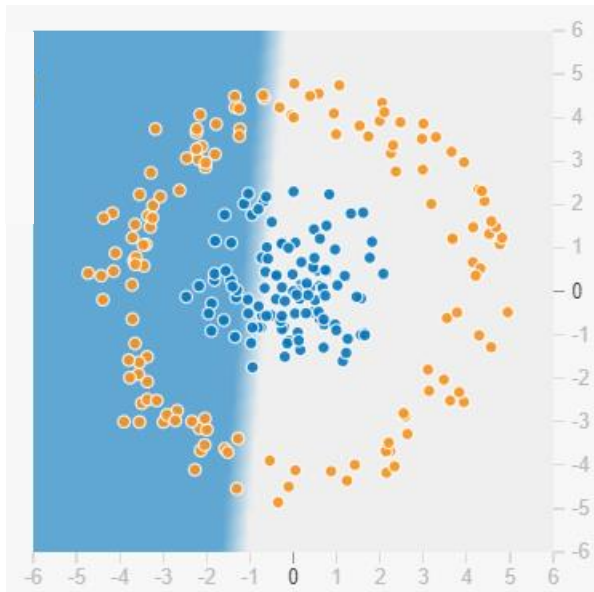
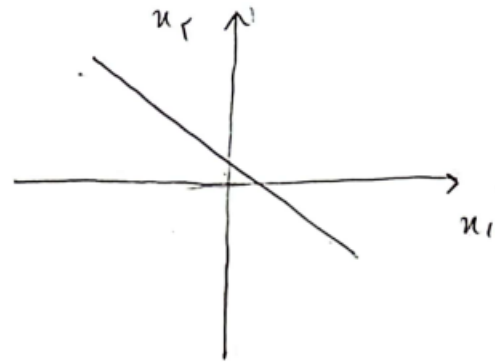


$$z_i: \gamma \forall q \quad x_i - 1, 2x_i + \gamma \cdot \omega \lambda = 0$$

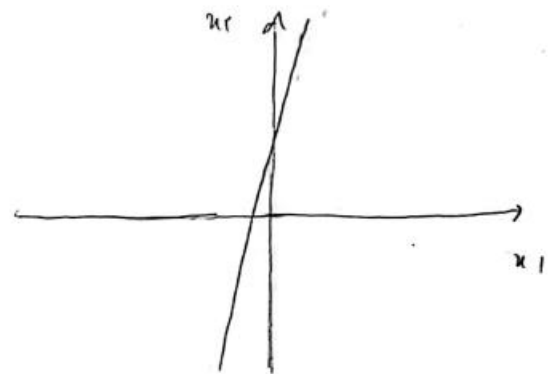


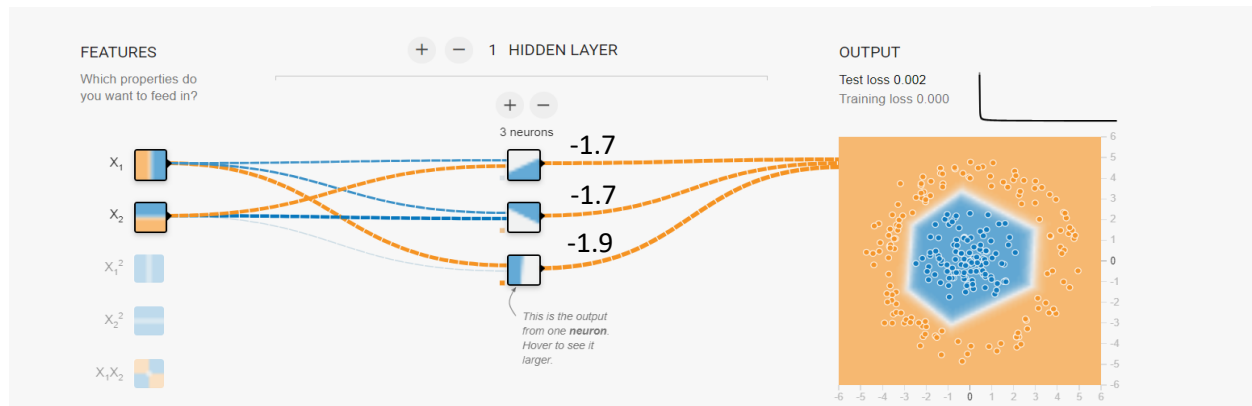


$$j \nearrow : 0.171 x_1 + 1.17 x_2 - 0.179 = 0$$



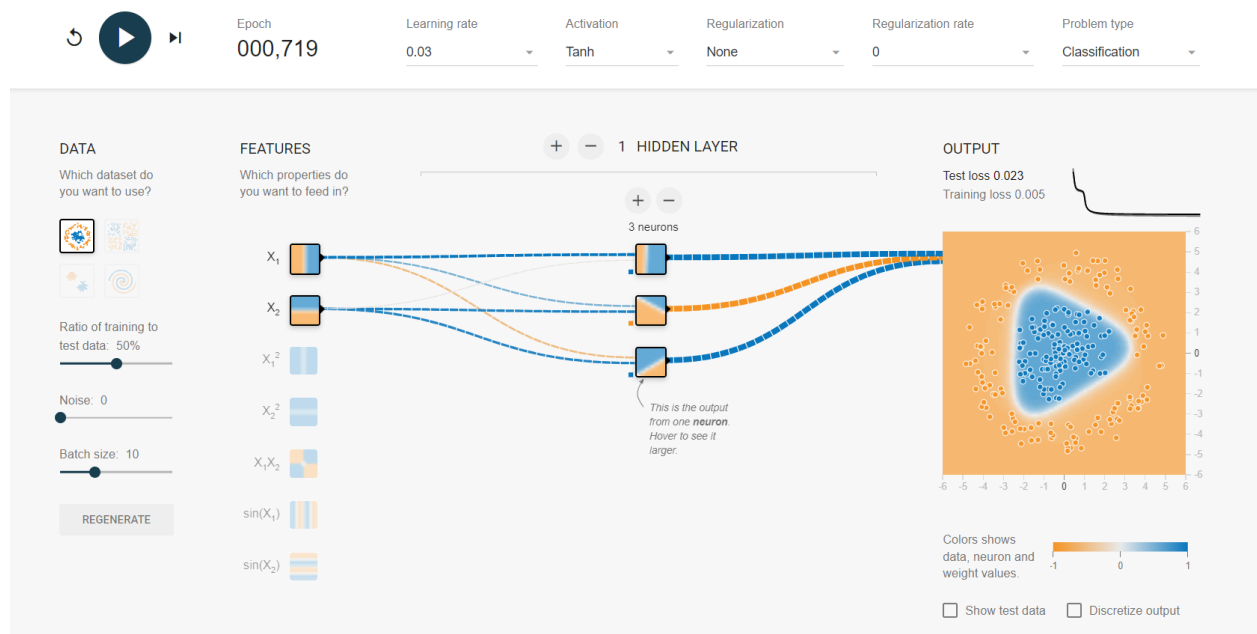
$$j \nearrow : -1.17 x_1 + 0.17 x_2 - 1.17 = 0$$

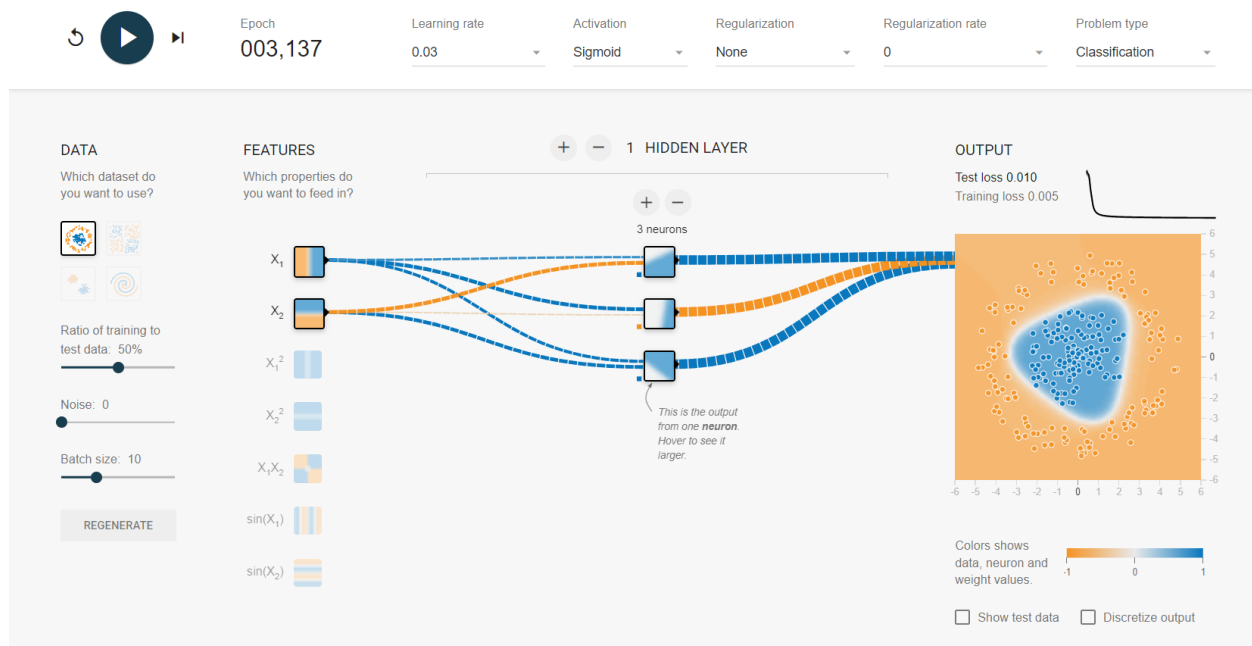




از آنجایی که محدوده تمام وزن‌های لایه خروجی حتی است، برای اینکه بتواند از محدوده‌ی مثبت را  
نه خروجی دریافت کنیم، باید عملیات با ریاض لایه خروجی مثبت باشد.  
زیرا، فقط لایه  $R$  به ازای هر وزن این اجزاء را می‌دهد تا پس از تبدیل سراسری مقادیر مثبتی‌های  
خطی با مدل سراسری کند که نتیجه بر این وزن‌های با خطوط مستقیم می‌شود.

(ج)





تفاوت اصلی در خطی بودن در مقابل غیر خطی بودن این توابع نهفته است. رفتار خطی ReLU بعد از صفر به آن اجازه می‌دهد تا پس از فعال‌سازی فقط وابستگی‌های خطی را مدل‌سازی کند، مگر اینکه از چندین لایه استفاده شود. این منجر به ایجاد مرزهایی با خطوط مستقیم می‌شود. Sigmoid و tanh، که ذاتاً غیرخطی هستند، می‌توانند مرزهای منحنی را حتی با یک لایه ایجاد کنند، که امکان پارتیشن‌بندی پیچیده‌تری از فضای ورودی را فراهم می‌کند.

## شبکه های عصبی پیچشی

### سوال اول

1- لایه های کانولوشنال فیلترهایی را در مناطق کوچک و محلی تصویر ورودی اعمال می کنند و به شبکه اجازه می دهند تا الگوهای محلی مانند لبه ها، بافت ها و اشکال ساده را ثبت کند. همانطور که شبکه عمیق تر می شود، این الگوهای محلی را می توان برای شناسایی ساختارهای پیچیده تر مانند اشیاء و چهره ها ترکیب کرد ولی لایه های کاملاً متصل در مقابل، هر نورون را در یک لایه به هر نورون در لایه بعدی متصل می کنند، بدون توجه به ساختار فضایی ورودی. این بدان معنی است که اطلاعات مکانی و موقعیت مکانی حفظ نمی شود و یادگیری سلسله مراتب مکانی را برای شبکه دشوار می کند.

2- لایه های کانولوشنال وزن هایی را در مکان های فضایی مختلف تصویر ورودی به اشتراک می گذارند و تعداد پارامترها را در مقایسه با لایه های کاملاً متصل به طور قابل توجهی کاهش می دهند. این به اشتراک گذاری پارامتر نه تنها بار محاسباتی را کاهش می دهد، بلکه به جلوگیری از **overfitting** بیش از حد کمک می کند، به خصوص در هنگام برخورد با داده های محدود ولی لایه های کاملاً متصل به تعداد زیادی پارامتر نیاز دارند زیرا هر نورون در یک لایه به هر نورون لایه بعدی متصل است. برای تصاویر بزرگ، این منجر به تعداد زیادی وزن می شود که منجر به هزینه های محاسباتی بالاتر و افزایش خطر **overfitting** می شود.

### سوال دوم

$$\begin{aligned}
 f &= [x, y, z] \Rightarrow \text{مکس } f = [z, y, x] \\
 \left. \begin{aligned}
 1z + 4y + 0x &= -2 \Rightarrow z + 4y = -2 \\
 4z + 0y + (-2)x &= -2 \Rightarrow 4z - 2x = -2 \\
 0z + (-2)y + 2x &= 11 \Rightarrow -2y + 2x = 11
 \end{aligned} \right\} \Rightarrow \begin{aligned}
 2x + 4y &= -4 \\
 4z - 2x &= -2 \\
 2x - 2y &= 11
 \end{aligned}
 \end{aligned}$$

$$\Rightarrow x = 3, 15 \Rightarrow y = -0.77 \Rightarrow z = 1.075$$

$$\Rightarrow f = [3, 15, -0.77, 1.075]$$