1 point	**1.** Consider a vector $(1, -2, 0.5)$. Apply a softmax transform to it and enter the first component (accurate to 2 decimal places).

0.60

1 point	**2.** Suppose you are solving a 5-class classification problem with 10 features. How many parameters a linear model would have? Don't forget bias terms!

55

1 point	**3.** There is an analytical solution for linear regression parameters and MSE loss, but we usually prefer gradient descent optimization over it. What are the reasons?

- ☑ Gradient descent doesn't require to invert a matrix

- ☑ Gradient descent is more scalable and can be applied for problems with high number of features

- ☐ Gradient descent is a method developed especially for MSE loss

- ☐ Gradient descent can find parameter values that give lower MSE value than parameters from analytical solution