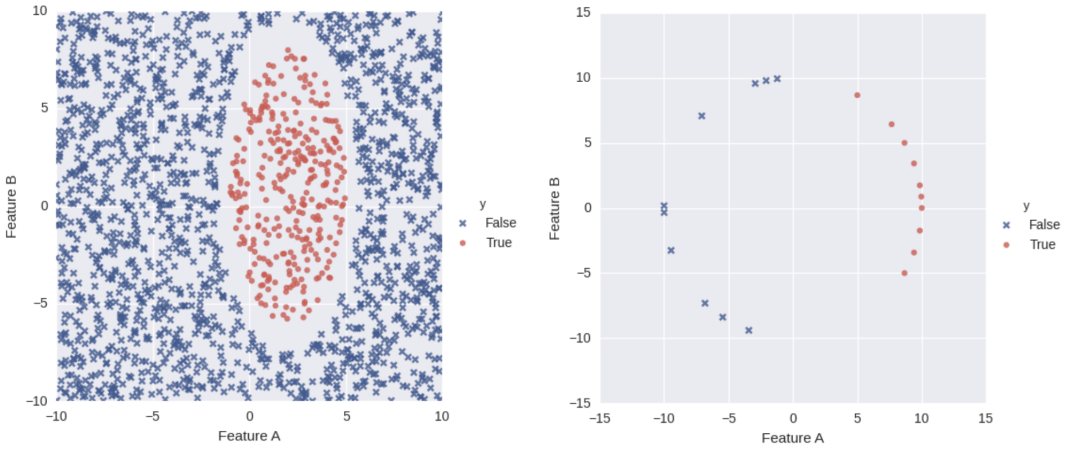
Here are some extra practice problems taken from past semesters. You are welcomed to check your solutions with me or discuss them in the telegram group chat.

1. (AY 24/25 Sem 1 Final Exam) You are minimizing the cost function $J(w) = \frac{1}{2}w^2 + 4w$ using gradient descent, what is the **largest** integer learning rate α that can be used so that it always finds the optimal value regardless of the initial value?

2. (AY 24/25 Sem 1 Final Exam) Consider a logistic regression model for multi-class classification with three classes: Pizza, Burger, and Sushi. The weight vectors for our multi-class (One vs One) classifiers where the $h_{\text{A/B}}(x)$ represents the probability of the class A. The weight vectors for each classifier include the bias term as the first element in each weight vector (2 is the bias for $w_{\text{Pizza/Burger}}$).
   $w_{\text{Pizza/Burger}} = \begin{bmatrix}2 & -0.5 & 0.3\end{bmatrix}$
   $w_{\text{Sushi/Pizza}} = \begin{bmatrix}-1 & 0.2 & -0.4\end{bmatrix}$
   $w_{\text{Burger/Sushi}} = \begin{bmatrix}0 & 0.4 & 0.1\end{bmatrix}$
   Given the input $\begin{bmatrix}3 & 2\end{bmatrix}$, determine which class the model predicts. Justify your answer.

3. For the following two cases, define a minimal set of features that will perfectly classify the data. Here are some examples: $(A), (B), (AB, A^{10})$.



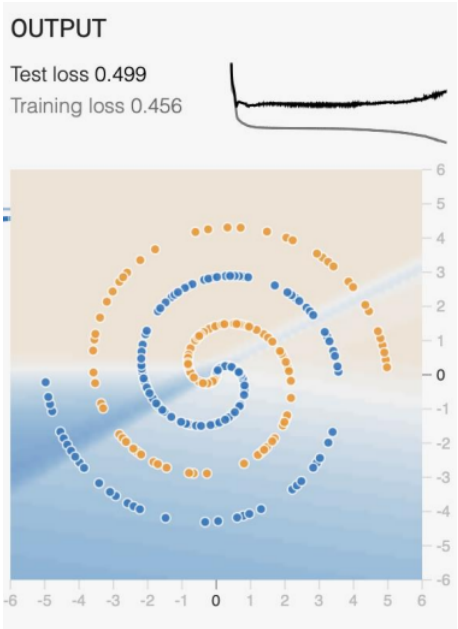
4. (AY 24/25 Sem 1 Final Exam) Suppose you are training a classifier with stochastic gradient descent.

   (a) Which evaluation metric is the **most** appropriate to evaluate the model's performance?

   A. Accuracy
   B. Precision and recall
   C. Mean squared error
   D. Weighted binary cross entropy loss

   (b) The result shown in the figure is after 1300 epochs with step size 0.03. Which advice(s) will you give to your colleague?

   A. The test loss is too high, so they should use more test examples.
   B. Keep running for more epochs.
   C. Use a less complex model.
   D. Use transformed features.



Answers

1. $\alpha = 1$
2. Probabilities: 0.75, 0.23, 0.80. Conclusion: Pizza.
3. Left: (A^2, B^2, A, B), Right: (A)
4. (a) B (b) BD or BCD