Framing: Check Your Understanding

Supervised Learning

Explore the options below.

Suppose you want to develop a supervised machine learning model to predict whether a given email is "spam" or "not spam." Which of the following statements are true?

Words in the subject header will make good labels. ⌄

We'll use unlabeled examples to train the model. ⌄

The labels applied to some examples might be unreliable. ✓

Definitely. It's important to check how reliable your data is. The labels for this dataset probably come from email users who mark particular email messages as spam. Since most users do not mark every suspicious email message as spam, we may have trouble knowing whether an email is spam. Furthermore, spammers could intentionally poison our model by providing faulty labels.

2 of 2 correct answers.

Emails not marked as "spam" or "not spam" are unlabeled examples. ✓
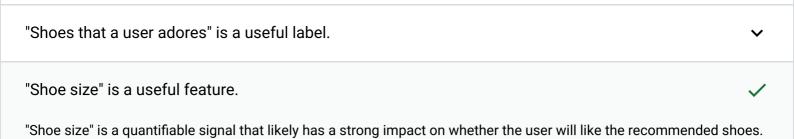
Because our label consists of the values "spam" and "not spam", any email not yet marked as spam or not spam is an unlabeled example.

1 of 2 correct answers.

Features and Labels

Explore the options below.

Suppose an online shoe store wants to create a supervised ML model that will provide personalized shoe recommendations to users. That is, the model will recommend certain pairs of shoes to Marty and different pairs of shoes to Janet. The system will use past user behavior data to generate training data. Which of the following statements are true?

"Shoe beauty" is a useful feature. ⌄

"Shoes that a user adores" is a useful label. ⌄

"Shoe size" is a useful feature. ✓

"Shoe size" is a quantifiable signal that likely has a strong impact on whether the user will like the recommended shoes. For example, if Marty wears size 9, the model shouldn't recommend size 7 shoes.

2 of 2 correct answers.

"The user clicked on the shoe's description" is a useful label. ✓

Users probably only want to read more about those shoes that they like. Clicks by users is, therefore, an observable, quantifiable metric that could serve as a good training label. Since our training data derives from past user behavior, our labels need to derive from objective behaviors like clicks that strongly correlate with user preferences.
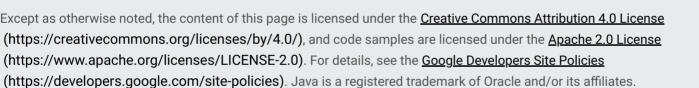
1 of 2 correct answers.

Help Center (https://support.google.com/machinelearningeducation)