

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?

- The labels applied to some examples might be untrustworthy.

Definitely. The labels for this dataset probably come from email users who mark particular email messages as spam. Since very few users mark every suspicious email message as spam, we may have a hard time ever knowing whether an email is spam. Furthermore, some spammers or botnets could intentionally poison our model by providing faulty labels.

1 of 2 correct answers.

- Words in the subject header will make good labels.
- 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.

2 of 2 correct answers.

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

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. Which of the following statements are true?

- The shoes that a user adores is a useful label.

- User clicks on a shoe's description is a useful label.

Users probably only want to read more about those shoes that they like. User clicks is, therefore, an observable, quantifiable metric that could serve as a good training label.

2 of 2 correct answers.

- Shoe beauty is a useful feature.
- 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.

1 of 2 correct answers.