

Filter

Help Center

ML models

- Linear regression (70 min)
- Logistic regression (35 min)
- Classification (70 min)
 - Introduction (3 mins)
 - Thresholds and the confusion matrix (12 min)
 - Accuracy, recall, precision, and related metrics (15 min)
 - ROC and AUC (10 min)
 - Prediction bias (3 min)
 - Multi-class classification (2 min)
- Programming exercise (15 min)
- Test your knowledge (10 min)
- What's next



Home > Products > Machine Learning > ML Concepts > Crash Course

Was this helpful?



Classification: Prediction bias



Send feedback

As mentioned in the [Linear regression](#) module, calculating **prediction bias** is a quick check that can flag issues with the model or training data early on.

Prediction bias is the difference between the mean of a model's **predictions** and the mean of **ground-truth** labels in the data. A model trained on a dataset where 5% of the emails are spam should predict, on average, that 5% of the emails it classifies are spam. In other words, the mean of the labels in the ground-truth dataset is 0.05, and the mean of the model's predictions should also be 0.05. If this is the case, the model has zero prediction bias. Of course, the model might still have other problems.

If the model instead predicts 50% of the time that an email is spam, then something is wrong with the training dataset, the new dataset the model is applied to, or with the model itself. Any significant difference between the two means suggests that the model has some prediction bias.

Prediction bias can be caused by:

- Biases or noise in the data, including biased sampling for the training set
- Too-strong regularization, meaning that the model was oversimplified and lost some necessary complexity
- Bugs in the model training pipeline
- The set of features provided to the model being insufficient for the task



Key terms:

- [Ground truth](#)
- [Prediction](#)
- [Prediction bias](#)

Help Center

Previous



ROC and AUC (10 min)

Next

Multi-class classification (2 min)



Was this helpful?



Send feedback

Except as otherwise noted, the content of this page is licensed under the [Creative Commons Attribution 4.0 License](#), and code samples are licensed under the [Apache 2.0 License](#). For details, see the [Google Developers Site Policies](#). Java is a registered trademark of Oracle and/or its affiliates.

Last updated 2024-10-09 UTC.

Connect

Blog

Instagram

LinkedIn

X (Twitter)

YouTube

Programs

Google Developer Groups

Google Developer Experts

Accelerators

Women Techmakers

Google Cloud & NVIDIA

Developer consoles

Google API Console

Google Cloud Platform Console

Google Play Console

Firebase Console

Actions on Google Console

Cast SDK Developer Console

Chrome Web Store Dashboard

Google Home Developer Console

Google for Developers

Android

Chrome

Firebase

Google Cloud Platform

Google AI

All products

Terms | Privacy

Sign up for the Google for Developers newsletter

Subscribe

English