Difference Between Classification and Regression in Machine Learning

by **Jason Brownlee** on December 11, 2017 in **Start Machine Learning**

There is an important difference between classification and regression problems.

Fundamentally, classification is about predicting a label and regression is about predicting a quantity.

I often see questions such as:

> *How do I calculate accuracy for my regression problem?*

Questions like this are a symptom of not truly understanding the difference between classification and regression and what accuracy is trying to measure.

In this tutorial, you will discover the differences between classification and regression.

After completing this tutorial, you will know:

- That predictive modeling is about the problem of learning a mapping function from inputs to outputs called function approximation.
- That classification is the problem of predicting a discrete class label output for an example.
- That regression is the problem of predicting a continuous quantity output for an example.

Let's get started.

Difference Between Classification and Regression in Machine Learning
Photo by thomas wilson, some rights reserved.

Tutorial Overview

This tutorial is divided into 5 parts; they are:

1. Function Approximation
2. Classification
3. Regression
4. Classification vs Regression

5. Converting Between Classification and Regression Problems

Function Approximation

Predictive modeling is the problem of developing a model using historical data to make a prediction on new data where we do not have the answer.

For more on predictive modeling, see the post:

- Gentle Introduction to Predictive Modeling

Predictive modeling can be described as the mathematical problem of approximating a mapping function (f) from input variables (X) to output variables (y). This is called the problem of function approximation.

The job of the modeling algorithm is to find the best mapping function we can given the time and resources available.

For more on approximating functions in applied machine learning, see the post:

- How Machine Learning Algorithms Work

Generally, we can divide all function approximation tasks into classification tasks and regression tasks.

Classification Predictive Modeling

Classification predictive modeling is the task of approximating a mapping function (f) from input variables (X) to discrete output variables (y).

The output variables are often called labels or categories. The mapping function predicts the class or category for a given observation.

For example, an email of text can be classified as belonging to one of two classes: “spam“ and “*not spam*“.

- A classification problem requires that examples be classified into one of two or more classes.
- A classification can have real-valued or discrete input variables.
- A problem with two classes is often called a two-class or binary classification problem.
- A problem with more than two classes is often called a multi-class classification problem.

- A problem where an example is assigned multiple classes is called a multi-label classification problem.

It is common for classification models to predict a continuous value as the probability of a given example belonging to each output class. The probabilities can be interpreted as the likelihood or confidence of a given example belonging to each class. A predicted probability can be converted into a class value by selecting the class label that has the highest probability.

For example, a specific email of text may be assigned the probabilities of 0.1 as being “spam” and 0.9 as being “not spam”. We can convert these probabilities to a class label by selecting the “not spam” label as it has the highest predicted likelihood.

There are many ways to estimate the skill of a classification predictive model, but perhaps the most common is to calculate the classification accuracy.

The classification accuracy is the percentage of correctly classified examples out of all predictions made.

For example, if a classification predictive model made 5 predictions and 3 of them were correct and 2 of them were incorrect, then the classification accuracy of the model based on just these predictions would be:

```
accuracy = correct predictions / total predictions * 100
accuracy = 3 / 5 * 100
accuracy = 60%
```

An algorithm that is capable of learning a classification predictive model is called a classification algorithm.

Regression Predictive Modeling

Regression predictive modeling is the task of approximating a mapping function (f) from input variables (X) to a continuous output variable (y).

A continuous output variable is a real-value, such as an integer or floating point value. These are often quantities, such as amounts and sizes.

For example, a house may be predicted to sell for a specific dollar value, perhaps in the range of $100,000 to $200,000.

- A regression problem requires the prediction of a quantity.
- A regression can have real valued or discrete input variables.
- A problem with multiple input variables is often called a multivariate regression problem.

- A regression problem where input variables are ordered by time is called a time series forecasting problem.

Because a regression predictive model predicts a quantity, the skill of the model must be reported as an error in those predictions.

There are many ways to estimate the skill of a regression predictive model, but perhaps the most common is to calculate the root mean squared error, abbreviated by the acronym RMSE.

For example, if a regression predictive model made 2 predictions, one of 1.5 where the expected value is 1.0 and another of 3.3 and the expected value is 3.0, then the RMSE would be:

```
RMSE = sqrt(average(error^2))
RMSE = sqrt(((1.0 - 1.5)^2 + (3.0 - 3.3)^2) / 2)
RMSE = sqrt((0.25 + 0.09) / 2)
RMSE = sqrt(0.17)
RMSE = 0.412
```

A benefit of RMSE is that the units of the error score are in the same units as the predicted value.

An algorithm that is capable of learning a regression predictive model is called a regression algorithm.

Some algorithms have the word “regression” in their name, such as linear regression and logistic regression, which can make things confusing because linear regression is a regression algorithm whereas logistic regression is a classification algorithm.

Classification vs Regression

Classification predictive modeling problems are different from regression predictive modeling problems.

- Classification is the task of predicting a discrete class label.
- Regression is the task of predicting a continuous quantity.

There is some overlap between the algorithms for classification and regression; for example:

- A classification algorithm may predict a continuous value, but the continuous value is in the form of a probability for a class label.
- A regression algorithm may predict a discrete value, but the discrete value in the form of an integer quantity.

Some algorithms can be used for both classification and regression with small modifications, such as decision trees and artificial neural networks. Some algorithms cannot, or cannot easily be used for both problem types, such as linear regression for regression predictive modeling and logistic regression for classification predictive modeling.

Importantly, the way that we evaluate classification and regression predictions varies and does not overlap, for example:

- Classification predictions can be evaluated using accuracy, whereas regression predictions cannot.
- Regression predictions can be evaluated using root mean squared error, whereas classification predictions cannot.

Convert Between Classification and Regression Problems

In some cases, it is possible to convert a regression problem to a classification problem. For example, the quantity to be predicted could be converted into discrete buckets.

For example, amounts in a continuous range between $0 and $100 could be converted into 2 buckets:

- Class 0: $0 to $49
- Class 1: $50 to $100

This is often called discretization and the resulting output variable is a classification where the labels have an ordered relationship (called ordinal).

In some cases, a classification problem can be converted to a regression problem. For example, a label can be converted into a continuous range.

Some algorithms do this already by predicting a probability for each class that in turn could be scaled to a specific range:

```
quantity = min + probability * range
```

Alternately, class values can be ordered and mapped to a continuous range:

- $0 to $49 for Class 1
- $50 to $100 for Class 2

If the class labels in the classification problem do not have a natural ordinal relationship, the conversion from classification to regression may result in surprising or poor performance as the model may learn a false or non-existent mapping from inputs to the continuous output range.

Further Reading

This section provides more resources on the topic if you are looking to go deeper.

- Gentle Introduction to Predictive Modeling
- How Machine Learning Algorithms Work

Summary

In this tutorial, you discovered the difference between classification and regression problems.

Specifically, you learned:

- That predictive modeling is about the problem of learning a mapping function from inputs to outputs called function approximation.
- That classification is the problem of predicting a discrete class label output for an example.
- That regression is the problem of predicting a continuous quantity output for an example.

Do you have any questions?
Ask your questions in the comments below and I will do my best to answer.

About Jason Brownlee

Jason Brownlee, PhD is a machine learning specialist who teaches developers how to get results with modern machine learning methods via hands-on tutorials.

View all posts by Jason Brownlee →

119 Responses to *Difference Between Classification and Regression in Machine Learning*

Rizwan ali December 11, 2017 at 7:14 am #

REPLY

nice post to clear the basic concepts

Jason Brownlee December 11, 2017 at 4:50 pm #

REPLY

Thanks!

shivaprasad December 11, 2017 at 6:47 pm #

REPLY

Really a good one sir,i had little bit confusion regarding these two.Thank you very much

Jason Brownlee December 12, 2017 at 5:27 am #

REPLY

I'm glad it helped.

Kishore December 11, 2017 at 8:54 pm #

REPLY

Very specific and crystal clear.. we need more of these in your books too 🙂

Jason Brownlee December 12, 2017 at 5:28 am #

REPLY

Thanks.

James December 13, 2017 at 12:02 am #

REPLY

“Some algorithms have the word “regression” in their name, such as linear regression and logistic regression, which can make things confusing because linear regression is a regression algorithm whereas logistic regression is a classification algorithm”

Thank you for that clarification, even after all this time in ML that always nagged at me.\

Jason Brownlee December 13, 2017 at 5:38 am #

REPLY

I’m glad it helped James.

Andrey Koch December 15, 2017 at 5:36 am #

REPLY

This topic is known to me yet the article made it straight to the point, thanks!

One thing, in “A classification can have real-valued or discrete input variables.” you probably meant continious vs. discrete

Jason Brownlee December 15, 2017 at 5:40 am #

REPLY

Thanks.

Derek Amah December 15, 2017 at 5:44 am #

REPLY

Very basic and insightful. I experience same misunderstanding from people when hiring talent and sometimes explaining these concepts to people. Will take a cue from your explanation going forward. Thanks !!!!

Monsef December 15, 2017 at 6:12 am #

REPLY

Valuable tutorial Jason

Thank you very much

Jason Brownlee December 15, 2017 at 3:30 pm #

REPLY

hanks, I'm glad it helped.

youssef December 15, 2017 at 8:25 am #

REPLY

Merci bien pour votre clarification de la différences entre les deux types ..

Jason Brownlee December 15, 2017 at 3:32 pm #

REPLY

You're welcome, I'm glad it helped.

Manika.rao December 15, 2017 at 4:15 pm #

REPLY

Thank a lot Sir.its indeed a good post for basics on ml….

Jason Brownlee December 16, 2017 at 5:23 am #

REPLY

I'm glad it helped.

JP December 16, 2017 at 3:20 am #

REPLY

Hey Jason,

This is really helpful for beginners.
Could you provide single example of following section:
"Convert Between Classification and Regression Problems"

Jason Brownlee December 16, 2017 at 5:34 am #

REPLY

Thanks for the suggestion.

MJ December 18, 2017 at 7:13 am #

REPLY

Great explanation of important concepts. Liked your comment on logistic regresión being a classification problem. The regression in the name does make it confusing when starting out.

Jason Brownlee December 18, 2017 at 3:22 pm #

REPLY

I'm glad it helped!

Idrees December 18, 2017 at 5:32 pm #

REPLY

I'm actually getting to focus on ML, absolutely new to it. Thanks for your eye-opening and insightful explanations

Jason Brownlee December 19, 2017 at 5:15 am #

REPLY

I'm glad the post helped.

alejandro Camargo December 20, 2017 at 5:31 am #

REPLY

Thank you!

Jason Brownlee December 20, 2017 at 5:52 am #

REPLY

You're welcome.

Mark December 29, 2017 at 8:14 am #

REPLY

How is it possible to reverse the process of discretizing data given that there is a loss of information? For example, following the same example above values of $0 to $49 would be represented by categorical value of 0.

Can't figure out how the reverse is possible without knowing original value (in this case price in USD).

Jason Brownlee December 29, 2017 at 2:35 pm #

REPLY

You cannot as information is lost.

Kamran January 1, 2018 at 8:45 am #

REPLY

Any help about CNN and RNN implementation using pima indian dataset in python

Jason Brownlee January 2, 2018 at 5:31 am #

REPLY

The Pima Indians dataset would be a bad for for a CNN+LSTM model.

You can learn about the model here:
https://machinelearningmastery.com/cnn-long-short-term-memory-networks/

JOHN JEFFRY MENDEZ January 22, 2018 at 8:01 pm #

REPLY

Great blog! could u please explain how to determine the variance and bias for a model prediction?

Jason Brownlee January 23, 2018 at 7:54 am #

REPLY

No, they are abstract concepts for understanding how algorithms work:
https://machinelearningmastery.com/gentle-introduction-to-the-bias-variance-trade-off-in-machine-learning/

doupanpan January 25, 2018 at 5:11 am #

REPLY

Thanks for the post, clearly explained. Also, I am thinking about using Python to try these classification. regression, models etc. is there a useful online tutorials to follow ?

Jason Brownlee January 25, 2018 at 5:58 am #

REPLY

Yes, right here:
https://machinelearningmastery.com/start-here/#python

Oussama January 25, 2018 at 7:37 am #

REPLY

Nice explanation. thanks.

Jason Brownlee January 25, 2018 at 9:10 am #

REPLY

You're welcome.

Dhanashree February 1, 2018 at 9:08 pm #

REPLY

Thanks very well explained. !!

Jason Brownlee February 2, 2018 at 8:18 am #

REPLY

Thanks.

Phil Mckay February 6, 2018 at 1:56 am #

REPLY

Hi Jason:

I am a patent attorney, used to be a physics animal. I have written several hundred patent applications directed to software, mostly cyber security and cloud security system patents. Lately, like in every field, AI and ML in particular is everywhere. I am not afraid of the math, but the vocabulary was an issue – until I read your post.

I can't tell you how much I appreciated that. As you know, vocabulary, at least consistent vocabulary usage, is a bit of an issue in the data science and software worlds. Your post cleared up many questions in just a few minutes of reading – THANK YOU!

You are an awesome writer and teacher

Jason Brownlee February 6, 2018 at 9:19 am #

REPLY

I'm glad it helped Phil!

Wafa February 9, 2018 at 4:04 am #

REPLY

Hi,

Thank you for your great tutorials.
Could you please give us a tutorial on how to classify images using transfer learning and Tensorflow?
Or guide me where I can find great tutorial about it like yours?

REPLY

Jason Brownlee February 9, 2018 at 9:15 am #

I have an example in Keras here:
https://machinelearningmastery.com/use-pre-trained-vgg-model-classify-objects-photographs/

REPLY

Wafa February 10, 2018 at 9:33 am #

Wow! thank you for replying

REPLY

Ali Shan February 21, 2018 at 6:05 am #

Appreciate your effort and clearly helpful.
thanks !

REPLY

Jason Brownlee February 21, 2018 at 6:42 am #

You're welcome.

hans March 5, 2018 at 6:39 am #

REPLY

simple and valuable , thanks.

Jason Brownlee March 6, 2018 at 6:06 am #

REPLY

Thanks, I'm glad to hear that.

KK March 15, 2018 at 6:06 pm #

REPLY

very well explained, actually the solution of my problem.
"How do I calculate accuracy for my regression problem?" this was my actual question but now it's clear.
Thanks man !!

Jason Brownlee March 16, 2018 at 6:10 am #

REPLY

I'm so glad to hear that!

Frederick Alfhendra March 30, 2018 at 2:58 am #

REPLY

Hello Sir, did you had any suggestion regarding the reference such as Journal or Book that I could use to explain about Regression Predictive Modeling?

Thank you

Jason Brownlee March 30, 2018 at 6:42 am #

REPLY

What questions do you have exactly?

Frederick Alfhendra March 30, 2018 at 12:50 pm #

REPLY

The question will be, “why we use Regression Predictive Modeling for Stock index forecasting”, any suggestion will be really appreciated, thank you, Sir.

Jason Brownlee March 31, 2018 at 6:32 am #

REPLY

I understand that short term price movements are a random walk and that a persistence model will be the best that you can achieve: https://machinelearningmastery.com/gentle-introduction-random-walk-times-series-forecasting-python/

Frederick Alfhendra March 31, 2018 at 11:27 am #

REPLY

thank you so much, Sir, I’m writing my bachelor thesis and I’ve read a lot of your research on google scholar, those are very helpful!

Jason Brownlee April 1, 2018 at 5:40 am #

REPLY

Thanks.

Abhijeet April 1, 2018 at 4:57 pm #

REPLY

Thanks Jason for amazing article…you Rock Jason !!!

Jason Brownlee April 2, 2018 at 5:21 am #

REPLY

Thanks, I'm glad it helped.

disouja April 12, 2018 at 7:38 pm #

REPLY

please clear with example that what is called classification and what is regression problem. how to get to know the problem is a classification problem or a regression problem.

Jason Brownlee April 13, 2018 at 6:39 am #

REPLY

Classification is about predicting a label (e.g. 'red'). Regression is about predicting a quantity (e.g. 100).

Does that help?

Julio Lee April 24, 2018 at 1:28 pm #

REPLY

Fantastic post, thank you for sharing! I was recently training a model as a binary classification problem using sigmoid as a single output. However, I was able to get far better results using MSE rather than binary cross-entropy. Since MSE is mostly used for regression, does this mean I was forced to convert it to a regression problem? This has been tickling my brain for quite some time now….

Jason Brownlee April 24, 2018 at 2:51 pm #

REPLY

Thanks.

MSE with a sigmoid output function? Wow, and it worked without error?

Be careful when evaluating the skill of the model, ensure it is doing what you think.

Rupesh May 2, 2018 at 4:22 pm #

REPLY

Thanks……

Jason Brownlee May 3, 2018 at 6:31 am #

REPLY

You're welcome.

Danilo May 9, 2018 at 3:25 am #

REPLY

Excellent post! I appreciate too much your work Ph.D. Jason Brownlee.
I have a question. What type of task should I perform if my dependent variable observations are dichotomous, but I need to infer continuous values?

Jason Brownlee May 9, 2018 at 6:27 am #

REPLY

If you have categorical inputs and require a real-valued output, this sounds like regression, and a challenging case.

Perhaps you can try modeling it directly as a regression problem and see how you go. You may want to integer encode or one hot encode the inputs.

Danilo May 9, 2018 at 7:25 am #

REPLY

Thank you very much for your help!
I have used two variants to determine my continuous output:
1. The probability of belonging to the positive class in a classification model.
2. The output of a regression model.
The accuracy in terms of a CMC curve for the classification model outperformed the regression model. But I am not sure if I have misunderstood some results.
Is it feasible to assume the probability of belonging to the positive class in a classification model as the similarity to this class? When the classifier outputs the probability (p) to belong to the negative class, I computed the probability to belong to the positive class as 1-p.

Jason Brownlee May 9, 2018 at 2:55 pm #

REPLY

Predicting class probability is a classification problem. Some algorithms can predict a probability.

Danilo May 9, 2018 at 3:17 pm #

Thank you very much, Dr. Jason Brownlee! Your advice is helpful for my work. You are very gentle sharing your great knowledge!!!!

Jason Brownlee May 10, 2018 at 6:26 am #

I'm glad it helped.

Jefferson Sankara June 6, 2018 at 5:54 am #

REPLY

Dear Jason,

The article really explained the concepts well. Thank you for the links to tutorials using Python.

Jason Brownlee June 6, 2018 at 6:42 am #

REPLY

Thanks, I'm glad it helped.

damak June 15, 2018 at 11:00 pm #

REPLY

This is simply awesome and it helped clarify a number of issues in ML. Thanks Dr. Brownlee

Jason Brownlee June 16, 2018 at 7:27 am #

REPLY

I'm glad it helped.

Dhanushka Sanjaya June 16, 2018 at 12:49 am #

REPLY

thanks for this article. it is very helpfull

Jason Brownlee June 16, 2018 at 7:28 am #

REPLY

I'm glad it helped.

alejo July 9, 2018 at 5:07 am #

REPLY

Hi, Jason.

What happens when a problem has many ordinal categories? For example, if I wanted to predict, I don't know, the goals (soccer) a team scores in a match (which generally will be in the range of 0 to 10) is a classification problem or a regression problem?

Jason Brownlee July 9, 2018 at 6:37 am #

REPLY

There are no rules. Try both and see which results in the better or more stable model.

James Y July 18, 2018 at 4:48 am #

REPLY

A basic question: are the classifiers generally faster than the regressors because the discrete outputs are easier to compute?

Jason Brownlee July 18, 2018 at 6:38 am #

REPLY

No.

Shailendra July 19, 2018 at 2:55 pm #

REPLY

Thanks a lot for sharing this beautiful concept, i am looking for more.

Jason Brownlee July 20, 2018 at 5:49 am #

REPLY

I'm glad it helped.

Vinit Tanna July 20, 2018 at 8:16 am #

REPLY

Really helpful.
Now my concept is crystal clear!!
Thanks 🙂

Jason Brownlee July 21, 2018 at 6:26 am #

REPLY

I'm glad to hear that.

william July 29, 2018 at 1:37 am #

REPLY

this is great, thanks again jason!

is there a certain threshold of a RMSE for a regression problem that would provide the impetus to discretize the output variable into an ordinal classification problem? i.e. 0.5?

i could be completely wrong here, but it seems like discretizing the output of a regression is simply just artificially inflating your predictive model's precision accuracy…

what are you thoughts?

Jason Brownlee July 29, 2018 at 6:13 am #

REPLY

I think it depends on the problem, whether the interpretation of the output as discrete quantities makes sense.

Brandon August 11, 2018 at 1:28 am #

REPLY

Thanks

Jason Brownlee August 11, 2018 at 6:13 am #

REPLY

You're welcome.

Rawan September 13, 2018 at 2:54 pm #

REPLY

Thanks Jason! Your language in explaining the topic was easy to understand.

Jason Brownlee September 14, 2018 at 6:33 am #

REPLY

Thanks, I'm happy that it helped.

Urise September 20, 2018 at 5:39 pm #

REPLY

Thanks Jason, it is great help to me.

One thing I need to confirm, if I use movielens dataset to predict the rating, the regression prediction is the correct one, right? What if I use classification prediction in this case? May I got wrong prediction?

Jason Brownlee September 21, 2018 at 6:25 am #

REPLY

You can choose to model it as regression or classification.

There is no objective “correct” in machine learning, there is just a dataset and your goals/stakeholders.

Jani October 1, 2018 at 4:06 am #

REPLY

Are there any particular scenario that you can mention like a question which asks to implement Linear Regression and another question which ask to implement logistic regression (classification) so that we can implement in R.

Jason Brownlee October 1, 2018 at 6:31 am #

REPLY

When predicting a quantity, use linear regression.

When predicting a binary label, use logistic regression.

JG October 2, 2018 at 1:45 am #

REPLY

Hi jason,
thanks.

I consider this tutorial more convenient for my question about Regression vs Classification (classes).

In the case you can convert your model from Regression to Classification, as it is explained in your last section, which could be more convenient to apply Regression (continuos) or Classification (discrete labels e.g. you divided your continuos output in some segment of interest and associated them to some artificial range such very low, low, medium, high, super, ultra, super ultra, tremendously, such it is the current case for frequency spectrum :-)))…Which model performs better Regression or Classification? using the same common layers (with the exception at the output layer and activation function?).

Jason Brownlee October 2, 2018 at 6:27 am #

REPLY

We cannot know which approach will perform best for a given problem, you must use systematic experimentation and discover what works best.

This is the job of applied machine learning.

JG October 2, 2018 at 8:09 am #

REPLY

OK . Thks

Data Scientist October 2, 2018 at 3:51 am #

REPLY

from content, you mentioned that A classification algorithm may predict a continuous value, but the continuous value is in the form of a probability for a class label.

from the summary, you mentioned That classification is the problem of predicting a discrete class label output for an example.

quite confusing

Jason Brownlee October 2, 2018 at 6:29 am #

REPLY

Sorry, perhaps ignore the prediction of a probability for now then and focus on predicting a real value for regression and a label for classification.

Anuja Deokar October 9, 2018 at 2:47 am #

REPLY

When it is suitable to use regression over classification?

Jason Brownlee October 9, 2018 at 8:46 am #

REPLY

Use regression when you want to predict a quantity.

Use classification when you want to predict a label.

Does that help?

Kotrappa October 16, 2018 at 9:47 pm #

REPLY

So nice discussion Thank you Jason , i am really interested in applications of Classification and Regression problems in real world like Automotive industry.

Jason Brownlee October 17, 2018 at 6:53 am #

REPLY

Thanks.

Farhad October 28, 2018 at 7:13 pm #

REPLY

Hi Jason – I want the input to be the name of the event e.g. “Swimming” and the output to be like “Physical:80%, Social:10%, Emotional:10%”.

Another Example:
“Cooking” is “Vocational:30, Social:20, Nutritional:50”

This is clearly not a classification problem, as the output is not binary. Do you agree it can be best implemented by regression models?

Thanks,

Jason Brownlee October 29, 2018 at 5:55 am #

REPLY

It is a multi-label classification task.

I don’t have examples of this type of task sorry.

Boo October 30, 2018 at 10:48 am #

REPLY

Hi Jason. I was hoping if you could help clarify. You say to use regression for predicting quantities and classification for labels. If I was interested in the predicting the probability of something being in a specific class, would it still be a classification problem?

Jason Brownlee October 30, 2018 at 2:11 pm #

REPLY

Yes, that is still classification.

Vipin Chauhan November 9, 2018 at 9:34 pm #

REPLY

You are always Awesome, Jason. Thanks for all your efforts. It cleared my doubt about these 2 ML methods.

Jason Brownlee November 10, 2018 at 6:06 am #

REPLY

I'm happy it helped.

Praveena Chandra November 12, 2018 at 2:01 pm #

REPLY

Hi Jason,
Thanks for the very informative blog. I do have a question. i want to predict Y, which can take the values 0,1or 2 based on an independent variable (X) which takes a value from 0 to 9. However, there are additional 8 categorical variables that are my control variables. I want to know if X predicts Y. I performed a ordinal logistic regression, however, I was told classification would be a better approach. My question is, how does one perform classification if control variables are to be taken into account too?

Jason Brownlee November 12, 2018 at 2:11 pm #

REPLY

Perhaps a model can take the dependent and control variables as input and predict y?

Also, perhaps try a one hot encoding if the vars are ordinal/categorical.

Max L. November 14, 2018 at 7:48 pm #

REPLY

Hi Jason,

is it possible to generalize the amount of data that is needed for a regression and classification Problem?
In my Opinion, to make a valid answer of how much something is, it needs more Data to learn from than "just" saying it is 1 or 0. Or is there no difference in the needed amount of Data for Regression/Classification Prolems?
Thanks for an answer!

Jason Brownlee November 15, 2018 at 5:30 am #

REPLY

The amount of data required depends on the complexity of the problem and the chosen algorithm.

Perhaps this will help:
https://machinelearningmastery.com/much-training-data-required-machine-learning/

Lolo December 19, 2018 at 3:37 am #

REPLY

Thank you for that great article. Back to how you start: “How do I calculate accuracy for my regression problem?” I am still not sure why this isn’t a valid question to ask. If your regression model predicts an amount between $0 and $100, we could still use various metrics to assess the accuracy of the prediction (average % error, max % error, error variance, etc.) No?

Jason Brownlee December 19, 2018 at 6:38 am #

REPLY

No. You calculate the error in the prediction for regression problems.

You can transform the regression into a classification problem, predict labels (e.g. at $10 increments) and report accuracy of that, but, it may not be the best way to model the problem (e.g. harder).

nikita December 19, 2018 at 6:34 am #

REPLY

This explanation is just too awesome man.I had already studied about these concepts in depth a month ago.But most of the concepts used to appear so difficult to me until I read ur post.I am about to start a ml project in a couple of days,and I can’t express how nicely ur post has cleared my concepts.Thank you so much

Jason Brownlee December 19, 2018 at 6:43 am #

REPLY

I'm happy it helped!

Kp December 27, 2018 at 4:45 am #

REPLY

Brilliant explanation of confusing topics

Jason Brownlee December 27, 2018 at 5:45 am #

REPLY

Thanks. I'm happy that it helped.

Abhishek December 29, 2018 at 12:26 am #

REPLY

why the output of the regression problem is called continuous?

Jason Brownlee December 29, 2018 at 5:53 am #

REPLY

It is a real valued number, e.g. fractional or no interruption – continuous.

Leave a Reply

Name (required)

Email (will not be published) (required)

Website

SUBMIT COMMENT

Welcome to Machine Learning Mastery!

Hi, I'm Jason Brownlee, PhD
I write tutorials to help developers (*like you*) get results with machine learning.

Read More

Want more help?

POPULAR

How to Develop a Neural Machine Translation System from Scratch
JANUARY 10, 2018

So, You are Working on a Machine Learning Problem…
APRIL 4, 2018

How to Develop an N-gram Multichannel Convolutional Neural Network for Sentiment Analysis
JANUARY 12, 2018

How to Make Predictions with Keras
APRIL 9, 2018

11 Classical Time Series Forecasting Methods in Python (Cheat Sheet)
AUGUST 6, 2018

You're Doing it Wrong. Why Machine Learning Does Not Have to Be So Hard
MARCH 28, 2018

How to Develop LSTM Models for Multi-Step Time Series Forecasting of Household Power Consumption
OCTOBER 10, 2018

You might also like…

- How to Install Python for Machine Learning
- Your First Machine Learning Project in Python
- Your First Neural Network in Python
- Your First Classifier in Weka
- Your First Time Series Forecasting Project
