Machine Learning Glossary

Definitions of common machine learning terms.

Accuracy

Percentage of correct predictions made by the model.

Algorithm

A method, function, or series of instructions used to generate a machine learning model. Examples include linear regression, decision trees, support vector machines, and neural networks.

Attribute

A quality describing an observation (e.g. color, size, weight). In Excel terms, these are column headers.

Bias metric

What is the average difference between your predictions and the correct value for that observation?

- **Low bias** could mean every prediction is correct. It could also mean half of your predictions are above their actual values and half are below, in equal proportion, resulting in low average difference.
- **High bias** (with low variance) suggests your model may be underfitting and you're using the wrong architecture for the job.

Bias term

Allow models to represent patterns that do not pass through the origin. For example, if all my features were 0, would my output also be zero? Is it possible there is some base value upon which my features have an effect? Bias terms typically accompany weights and are attached to neurons or filters.

Categorical Variables

Variables with a discrete set of possible values. Can be ordinal (order matters) or nominal (order doesn't matter).

Classification

Predicting a categorical output (e.g. yes or no?, blue, green or red?).

Classification Threshold

The lowest probability value at which we're comfortable asserting a positive classification. For example, if the predicted probability of being diabetic is > 50%, return True, otherwise return False.

Clustering

Unsupervised grouping of data into buckets.

Confusion Matrix

Table that describes the performance of a classification model by grouping predictions into 4 categories.

- **True Positives**: we *correctly* predicted they do have diabetes
- **True Negatives**: we *correctly* predicted they don't have diabetes
- **False Positives**: we *incorrectly* predicted they do have diabetes (Type I error)
- **False Negatives**: we *incorrectly* predicted they don't have diabetes (Type II error)

Continuous Variables

Variables with a range of possible values defined by a number scale (e.g. sales, lifespan).

Deduction

A top-down approach to answering questions or solving problems. A logic technique that starts with a theory and tests that theory with observations to derive a conclusion. E.g. We suspect X, but we need to test our hypothesis before coming to any conclusions.

Deep Learning

Deep Learning is derived from one machine learning algorithm called perceptron or malti layer perceptron that gain more and more attention nowadays because of its success in different fields like, computer vision to signal processing and medical diagnosis to self-driving cars. As all other AI algorithms deep learning is from decades, but now today we have more and more data and cheap computing power that make this algorithm really powerful to achive state of the art acuracy.

in modern world this algorithm knowns as artificial neural network. deep learning is much more than traditional artificial neural network. but it was highly influenced by machine learning's neural network and perceptron network.

Dimension

Dimention for machine learning and data scientist is differ from physics, here Dimention of data means how much feature you have in you data ocean(data-set). e.g in case of object detection application, flatten image size and color channel(e.g 28*28*3) is a feature of the input set. in case of house price pridiction (maybe) house size is the data-set so we call it 1 dimentional data.

Epoch

An epoch describes the number of times the algorithm sees the entire data set.

Extrapolation

Making predictions outside the range of a dataset. E.g. My dog barks, so all dogs must bark. In machine learning we often run into trouble when we extrapolate outside the range of our training data.

Feature

With respect to a dataset, a feature represents an attribute and value combination. Color is an attribute. "Color is blue" is a feature. In Excel terms, features are similar to cells. The term feature has other definitions in different contexts.

Feature Selection

Feature selection is the process of selecting relevant features from a data-set for creating a Machine Learning model.

Feature Vector

A list of features describing an observation with multiple attributes. In Excel we call this a row.

Hyperparameters

Hyperparameters are higher-level properties of a model such as how fast it can learn (learning rate) or complexity of a model. The depth of trees in a Decision

Tree or number of hidden layers in a Neural Networks are examples of hyper parameters.

Induction

A bottoms-up approach to answering questions or solving problems. A logic technique that goes from observations to theory. E.g. We keep observing X, so we <i>infer</i> that Y must be True.

Instance

A data point, row, or sample in a dataset. Another term for observation.

Learning Rate

The size of the update steps to take during optimization loops like Gradient Descent. With a high learning rate we can cover more ground each step, but we risk overshooting the lowest point since the slope of the hill is constantly changing. With a very low learning rate, we can confidently move in the direction of the negative gradient since we are recalculating it so frequently. A low learning rate is more precise, but calculating the gradient is time-consuming, so it will take us a very long time to get to the bottom.

Loss

Loss = true_value(from data-set)- predicted value(from ML-model) The lower the loss, the better a model (unless the model has over-fitted to the training data). The loss is calculated on training and validation and its interperation is how well the model is doing for these two sets. Unlike accuracy, loss is not a percentage. It is a summation of the errors made for each example in training or validation sets.

Machine Learning

Mitchell (1997) provides a succinct definition: “A computer program is said to learn from experience E with respect to some

class of tasks T and performance measure P, if its performance at tasks in T , as measured by P, improves with experience E.In simple language machine learning is a field in which human made algorithms have an ability learn by itself or predict future for unseen data.

Model

A data structure that stores a representation of a dataset (weights and biases). Models are created/learned when you train an algorithm on a dataset.

Neural Networks

Contribute a definition!

Normalization

Restriction of the values of weights in regression to avoid overfitting and improving computation speed.

Null Accuracy

Baseline accuracy that can be acheived by always predicting the most frequent class ("B has the highest frequency, so lets guess B every time").

Observation

A data point, row, or sample in a dataset. Another term for instance.

Overfitting

Overfitting occurs when your model learns the training data too well and incorporates details and noise specific to your dataset. You can tell a model is overfitting when it performs great on your training/validation set, but poorly on your test set (or new real-world data).

Precision

In the context of binary classification (Yes/No), precision measures the model's performance at classifying positive observations (i.e. "Yes"). In other words, when a positive value is predicted, how often is the prediction correct? We could game this metric by only returning positive for the single observation we are most confident in.

$$P = TruePositives/(TruePositives + FalsePositives)$$

Recall

Also called sensitivity. In the context of binary classification (Yes/No), recall measures how "sensitive" the classifier is at detecting positive instances. In other words, for all the true observations in our sample, how many did we “catch.” We could game this metric by always classifying observations as positive.

$$R = TruePositives / (TruePositives + FalseNegatives)$$

Recall vs Precision

Say we are analyzing Brain scans and trying to predict whether a person has a tumor (True) or not (False). We feed it into our model and our model starts guessing.

- **Precision** is the % of True guesses that were actually correct! If we guess 1 image is True out of 100 images and that image is actually True, then our precision is 100%! Our results aren’t helpful however because we missed 10 brain tumors! We were super precise when we tried, but we didn’t try hard enough.
- **Recall**, or Sensitivity, provides another lens which with to view how good our model is. Again let’s say there are 100 images, 10 with brain tumors, and we correctly guessed 1 had a brain tumor. Precision is 100%, but recall is 10%. Perfect recall requires that we catch all 10 tumors!

Regression

Predicting a continuous output (e.g. price, sales).

Regularization

Contribute a definition!

Reinforcement Learning

Training a model to maximize a reward via iterative trial and error.

Segmentation

Contribute a definition!

Specificity

In the context of binary classification (Yes/No), specificity measures the model’s performance at classifying negative observations (i.e. “No”). In other words, when the correct label is negative, how often is the prediction correct? We could game this metric if we predict everything as negative.

$$S = TrueNegatives/(TrueNegatives + FalsePositives)$$

Supervised Learning

Training a model using a labeled dataset.

Test Set

A set of observations used at the end of model training and validation to assess the predictive power of your model. How generalizable is your model to unseen data?

Training Set

A set of observations used to generate machine learning models.

Transfer Learning

Contribute a definition!

Type 1 Error

False Positives. Consider a company optimizing hiring practices to reduce false positives in job offers. A type 1 error occurs when candidate seems good and they hire him, but he is actually bad.

Type 2 Error

False Negatives. The candidate was great but the company passed on him.

Underfitting

Underfitting occurs when your model over-generalizes and fails to incorporate relevant variations in your data that would give your model more predictive power. You can tell a model is underfitting when it performs poorly on both training and test sets.

Universal Approximation Theorem

A neural network with one hidden layer can approximate any continuous function but only for inputs in a specific range. If you train a network on inputs between -2 and 2, then it will work well for inputs in the same range, but you can’t expect it to generalize to other inputs without retraining the model or adding more hidden neurons.

Unsupervised Learning

Training a model to find patterns in an unlabeled dataset (e.g. clustering).

Validation Set

A set of observations used during model training to provide feedback on how well the current parameters generalize beyond the training set. If training error decreases but validation error increases, your model is likely overfitting and you should pause training.

Variance

How tightly packed are your predictions for a particular observation relative to each other?

- **Low variance** suggests your model is internally consistent, with predictions varying little from each other after every iteration.
- **High variance** (with low bias) suggests your model may be overfitting and reading too deeply into the noise found in every training set.

---

accuracy

The fraction of **predictions** that a **classification model** got right. In **multi-class classification**, accuracy is defined as follows:

Accuracy=Correct Predictions / Total Number Of Examples

In **binary classification**, accuracy has the following definition:

Accuracy=True Positives+True Negatives /Total Number Of Examples

See **true positive** and **true negative**.

activation function

A function (for example, **ReLU** or **sigmoid**) that takes in the weighted sum of all of the inputs from the previous layer and then generates and passes an output value (typically nonlinear) to the next layer.

artificial general intelligence

A non-human mechanism that demonstrates a *broad range* of problem solving, creativity, and adaptability. For example, a program demonstrating artificial general intelligence could translate text, compose symphonies, *and* excel at games that have not yet been invented.

artificial intelligence

A non-human program or model that can solve sophisticated tasks. For example, a program or model that translates text or a program or model that identifies diseases from radiologic images both exhibit artificial intelligence.

Formally, **machine learning** is a sub-field of artificial intelligence. However, in recent years, some organizations have begun using the terms *artificial intelligence* and *machine learning* interchangeably.

backpropagation

The primary algorithm for performing **gradient descent** on **neural networks**. First, the output values of each node are calculated (and cached) in a forward pass. Then, the **partial derivative** of the error with respect to each parameter is calculated in a backward pass through the graph.

bias (ethics/fairness)

#fairness

1. Stereotyping, prejudice or favoritism towards some things, people, or groups over others. These biases can affect collection and interpretation of data, the design of a system, and how users interact with a system. Forms of this type of bias include:

- **automation bias**
- **confirmation bias**
- **experimenter's bias**
- **group attribution bias**
- **implicit bias**
- **in-group bias**
- **out-group homogeneity bias**

2. Systematic error introduced by a sampling or reporting procedure. Forms of this type of bias include:

- **coverage bias**
- **non-response bias**
- **participation bias**
- **reporting bias**
- **sampling bias**
- **selection bias**

An intercept or offset from an origin. Bias (also known as the **bias term**) is referred to as *b* or w_0 in machine learning models. For example, bias is the *b* in the following formula:

$$y' = b + w_1x_1 + w_2x_2 + \ldots w_nx_n$$

categorical data

Features having a discrete set of possible values. For example, consider a categorical feature named `house style`, which has a discrete set of three possible values: `Tudor,` `ranch, colonial`. By representing `house style` as categorical data, the model can learn the separate impacts of `Tudor`, `ranch`, and `colonial` on house price.

Sometimes, values in the discrete set are mutually exclusive, and only one value can be applied to a given example. For example, a `car maker` categorical feature would probably permit only a single value (`Toyota`) per example. Other times, more than one value may be applicable. A single car could be painted more than one different color, so a `car color` categorical feature would likely permit a single example to have multiple values (for example, `red` and `white`).

Categorical features are sometimes called **discrete features**.

---

clustering

Grouping related **examples**, particularly during **unsupervised learning**. Once all the examples are grouped, a human can optionally supply meaning to each cluster.

---

continuous feature

A floating-point feature with an infinite range of possible values. Contrast with **discrete feature**.

DataFrame

A popular datatype for representing datasets in **pandas**. A DataFrame is analogous to a table. Each column of the DataFrame has a name (a header), and each row is identified by a number.

---

data set or dataset

A collection of **examples**.

ensemble

A merger of the predictions of multiple **models**. You can create an ensemble via one or more of the following:

- different initializations
- different **hyperparameters**
- different overall structure

epoch

A full training pass over the entire dataset such that each example has been seen once. Thus, an epoch represents `N`/**batch size** training **iterations**, where `N` is the total number of examples.

false negative (FN)

An example in which the model mistakenly predicted the **negative class**. For example, the model inferred that a particular email message was not spam (the negative class), but that email message actually was spam.

false positive (FP)

An example in which the model mistakenly predicted the **positive class**. For example, the model inferred that a particular email message was spam (the positive class), but that email message was actually not spam.

feature

An input variable used in making **predictions**.

feature engineering

The process of determining which **features** might be useful in training a model, and then converting raw data from log files and other sources into said features. In TensorFlow, feature engineering often means converting raw log file entries to **tf.Example** protocol buffers. See also tf.Transform.

Feature engineering is sometimes called **feature extraction**.

feature extraction

Overloaded term having either of the following definitions:

- Retrieving intermediate feature representations calculated by an **unsupervised** or pretrained model (for example, **hidden layer** values in a **neural network**) for use in another model as input.
- Synonym for **feature engineering**.

gradient

The vector of **partial derivatives** with respect to all of the independent variables. In machine learning, the gradient is the vector of partial derivatives of the model function. The gradient points in the direction of steepest ascent.

gradient descent

A technique to minimize **loss** by computing the gradients of loss with respect to the model's parameters, conditioned on training data. Informally, gradient descent iteratively adjusts parameters, gradually finding the best combination of **weights** and bias to minimize loss.

heuristic

A quick solution to a problem, which may or may not be the best solution. For example, "With a heuristic, we achieved 86% accuracy. When we switched to a deep neural network, accuracy went up to 98%."

hidden layer

A synthetic layer in a **neural network** between the **input layer** (that is, the features) and the **output layer** (the prediction). Hidden layers typically contain an **activation function** (such as **ReLU**) for training. A **deep neural network** contains more than one hidden layer.

- **Hypothesis in Science**: Provisional explanation that fits the evidence and can be confirmed or disproved.
- **Hypothesis in Statistics**: Probabilistic explanation about the presence of a relationship between observations.
- **Hypothesis in Machine Learning**: Candidate model that approximates a target function for mapping examples of inputs to outputs.

labeled example

An example that contains **features** and a **label**. In supervised training, models learn from labeled examples.

Mean Squared Error (MSE)

The average squared loss per example. MSE is calculated by dividing the **squared loss** by the number of **examples**. The values that **TensorFlow Playground** displays for "Training loss" and "Test loss" are MSE.

model

The representation of what a machine learning system has learned from the training data.

Momentum

A sophisticated gradient descent algorithm in which a learning step depends not only on the derivative in the current step, but also on the derivatives of the step(s) that immediately preceded it. Momentum involves computing an exponentially weighted moving average of the gradients over time, analogous to momentum in physics. Momentum sometimes prevents learning from getting stuck in local minima.

multi-class classification

Classification problems that distinguish among more than two classes. For example, there are approximately 128 species of maple trees, so a model that categorized maple tree species would be multi-class. Conversely, a model that divided emails into only two categories (*spam* and *not spam*) would be a **binary classification model**.

negative class

In **binary classification**, one class is termed positive and the other is termed negative. The positive class is the thing we're looking for and the negative class is the other possibility. For example, the negative class in a medical test might be "not tumor." The negative class in an email classifier might be "not spam." See also **positive class**.

neural network

A model that, taking inspiration from the brain, is composed of layers (at least one of which is **hidden**) consisting of simple connected units or **neurons** followed by nonlinearities.

neuron

A node in a **neural network**, typically taking in multiple input values and generating one output value. The neuron calculates the output value by applying an **activation function** (nonlinear transformation) to a weighted sum of input values.

node (neural network)

A **neuron** in a **hidden layer**.

noise

Broadly speaking, anything that obscures the signal in a dataset. Noise can be introduced into data in a variety of ways. For example:

- Human raters make mistakes in labeling.
- Humans and instruments mis-record or omit feature values.

one-hot encoding

A sparse vector in which:

- One element is set to 1.
- All other elements are set to 0.

One-hot encoding is commonly used to represent strings or identifiers that have a finite set of possible values. For example, suppose a given botany dataset chronicles 15,000 different species, each denoted with a unique string identifier. As part of feature engineering, you'll probably encode those string identifiers as one-hot vectors in which the vector has a size of 15,000.

outliers

Values distant from most other values. In machine learning, any of the following are outliers:

- **Weights** with high absolute values.
- Predicted values relatively far away from the actual values.
- Input data whose values are more than roughly 3 standard deviations from the mean.

Outliers often cause problems in model training. **Clipping** is one way of managing outliers.

output layer

The "final" layer of a neural network. The layer containing the answer(s).

overfitting

Creating a model that matches the **training data** so closely that the model fails to make correct predictions on new data.

parameter

A variable of a model that the machine learning system trains on its own. For example, **weights** are parameters whose values the machine learning system gradually learns through successive training iterations. Contrast with **hyperparameter**.

partial derivative

A derivative in which all but one of the variables is considered a constant. For example, the partial derivative of $f(x, y)$ with respect to x is the derivative of f considered as a function of x alone (that is, keeping y constant). The partial derivative of f with respect to x focuses only on how x is changing and ignores all other variables in the equation.

positive class

In **binary classification**, the two possible classes are labeled as positive and negative. The positive outcome is the thing we're testing for. (Admittedly, we're simultaneously testing for both outcomes, but play along.) For example, the positive class in a medical test might be "tumor." The positive class in an email classifier might be "spam."

precision

A metric for **classification models**. Precision identifies the frequency with which a model was correct when predicting the **positive class**. That is:

Precision=True Positives/(True Positives+False Positives)

Q-learning

#rl

In reinforcement learning, an algorithm that allows an **agent** to learn the optimal **Q-function** of a **Markov decision process** by applying the **Bellman equation**. The Markov decision process models an **environment**.

random forest

An ensemble approach to finding the **decision tree** that best fits the training data by creating many decision trees and then determining the "average" one. The "random" part of the term refers to building each of the decision trees from a random selection of features; the "forest" refers to the set of decision trees.

recall

A metric for **classification models** that answers the following question: Out of all the possible positive labels, how many did the model correctly identify? That is:

Recall=True Positives/True Positives+False Negatives

Rectified Linear Unit (ReLU)

An **activation function** with the following rules:

- If input is negative or zero, output is 0.
- If input is positive, output is equal to input.

recurrent neural network

A **neural network** that is intentionally run multiple times, where parts of each run feed into the next run. Specifically, hidden layers from the previous run provide part of the input to the same hidden layer in the next run. Recurrent neural networks are particularly useful for evaluating sequences, so that the hidden layers can learn from previous runs of the neural network on earlier parts of the sequence.

regression model

A type of model that outputs continuous (typically, floating-point) values. Compare with **classification models**, which output discrete values, such as "day lily" or "tiger lily."

reinforcement learning (RL)

#rl

A family of algorithms that learn an optimal **policy**, whose goal is to maximize **return** when interacting with an **environment**. For example, the ultimate reward of most games is victory. Reinforcement learning systems can become expert at playing complex games by evaluating sequences of previous game moves that ultimately led to wins and sequences that ultimately led to losses.

sigmoid function

A function that maps logistic or multinomial regression output (log odds) to probabilities, returning a value between 0 and 1. The sigmoid function has the following formula:

a:

$$y=1/1+e^{-\sigma}$$

where σ in **logistic regression** problems is simply:

where σ in **logistic regression** problems is simply:

$$\sigma=b+w1x1+w2x2+\ldots wnxn$$

In other words, the sigmoid function converts σ into a probability between 0 and 1.

temporal data

Data recorded at different points in time. For example, winter coat sales recorded for each day of the year would be temporal data.

training set

The subset of the dataset used to train a model.

underfitting

Producing a model with poor predictive ability because the model hasn't captured the complexity of the training data. Many problems can cause underfitting, including:

- Training on the wrong set of features.
- Training for too few epochs or at too low a learning rate.

- Training with too high a regularization rate.
- Providing too few hidden layers in a deep neural network.

---

validation

A process used, as part of **training**, to evaluate the quality of a **machine learning** model using the **validation set**. Because the validation set is disjoint from the training set, validation helps ensure that the model's performance generalizes beyond the training set.

---

validation set

A subset of the dataset—disjoint from the training set—used in **validation**.

---

weight

A coefficient for a **feature** in a linear model, or an edge in a deep network. The goal of training a linear model is to determine the ideal weight for each feature. If a weight is 0, then its corresponding feature does not contribute to the model.

Types of Machine Learning Techniques