≡ Navigation

Machine Learning Mastery
Making Developers Awesome at Machine Learning

Search...

# 14 Different Types of Learning in Machine Learning

by **Jason Brownlee** on November 11, 2019 in **Start Machine Learning**

Last Updated on November 11, 2019

Machine learning is a large field of study that overlaps with and inherits ideas from many related fields such as artificial intelligence.

The focus of the field is learning, that is, acquiring skills or knowledge from experience. Most commonly, this means synthesizing useful concepts from historical data.

As such, there are many different types of learning that you may encounter as a practitioner in the field of machine learning: from whole fields of study to specific techniques.

In this post, you will discover a gentle introduction to the different types of learning that you may encounter in the field of machine learning.

After reading this post, you will know:

- Fields of study, such as supervised, unsupervised, and reinforcement learning.
- Hybrid types of learning, such as semi-supervised and self-supervised learning.
- Broad techniques, such as active, online, and transfer learning.

Let’s get started.

Types of Learning in Machine Learning
Photo by Lenny K Photography, some rights reserved.

## Types of Learning

Given that the focus of the field of machine learning is "*learning*," there are many types that you may encounter as a practitioner.

Some types of learning describe whole subfields of study comprised of many different types of algorithms such as "*supervised learning*." Others describe powerful techniques that you can use on your projects, such as "*transfer learning*."

There are perhaps 14 types of learning that you must be familiar with as a machine learning practitioner; they are:

**Learning Problems**

- 1. Supervised Learning
- 2. Unsupervised Learning
- 3. Reinforcement Learning

**Hybrid Learning Problems**

- 4. Semi-Supervised Learning
- 5. Self-Supervised Learning
- 6. Multi-Instance Learning

**Statistical Inference**

- 7. Inductive Learning
- 8. Deductive Inference
- 9. Transductive Learning

**Learning Techniques**

- 10. Multi-Task Learning
- 11. Active Learning
- 12. Online Learning
- 13. Transfer Learning
- 14. Ensemble Learning

In the following sections, we will take a closer look at each in turn.

**Did I miss an important type of learning?**
Let me know in the comments below.

# Learning Problems

First, we will take a closer look at three main types of learning problems in machine learning: supervised, unsupervised, and reinforcement learning.

## 1. Supervised Learning

Supervised learning describes a class of problem that involves using a model to learn a mapping between input examples and the target variable.

> *Applications in which the training data comprises examples of the input vectors along with their corresponding target vectors are known as supervised learning problems.*

— Page 3, Pattern Recognition and Machine Learning, 2006.

Models are fit on training data comprised of inputs and outputs and used to make predictions on test sets where only the inputs are provided and the outputs from the model are compared to the withheld target variables and used to estimate the skill of the model.

> *Learning is a search through the space of possible hypotheses for one that will perform well, even on new examples beyond the training set. To measure the accuracy of a hypothesis we give it a test set of examples that are distinct from the training set.*

— Page 695, Artificial Intelligence: A Modern Approach, 3rd edition, 2015.

There are two main types of supervised learning problems: they are classification that involves predicting a class label and regression that involves predicting a numerical value.

- **Classification**: Supervised learning problem that involves predicting a class label.
- **Regression**: Supervised learning problem that involves predicting a numerical label.

Both classification and regression problems may have one or more input variables and input variables may be any data type, such as numerical or categorical.

An example of a classification problem would be the MNIST handwritten digits dataset where the inputs are images of handwritten digits (pixel data) and the output is a class label for what digit the image represents (numbers 0 to 9).

An example of a regression problem would be the Boston house prices dataset where the inputs are variables that describe a neighborhood and the output is a house price in dollars.

Some machine learning algorithms are described as "*supervised*" machine learning algorithms as they are designed for supervised machine learning problems. Popular examples include: decision trees, support vector machines, and many more.

> *Our goal is to find a useful approximation f(x) to the function f(x) that underlies the predictive relationship between the inputs and outputs*

— Page 28, The Elements of Statistical Learning: Data Mining, Inference, and Prediction, 2nd edition, 2016.

Algorithms are referred to as "*supervised*" because they learn by making predictions given examples of input data, and the models are supervised and corrected via an algorithm to better predict the expected target outputs in the training dataset.

> *The term supervised learning originates from the view of the target y being provided by an instructor or teacher who shows the machine learning system what to do.*

— Page 105, Deep Learning, 2016.

Some algorithms may be specifically designed for classification (such as logistic regression) or regression (such as linear regression) and some may be used for both types of problems with minor modifications (such as artificial neural networks).

## 2. Unsupervised Learning

Unsupervised learning describes a class of problems that involves using a model to describe or extract relationships in data.

Compared to supervised learning, unsupervised learning operates upon only the input data without outputs or target variables. As such, unsupervised learning does not have a teacher correcting the model, as in the case of supervised learning.

> *In unsupervised learning, there is no instructor or teacher, and the algorithm must learn to make sense of the data without this guide.*

— Page 105, Deep Learning, 2016.

There are many types of unsupervised learning, although there are two main problems that are often encountered by a practitioner: they are clustering that involves finding groups in the data and density estimation that involves summarizing the distribution of data.

- **Clustering: Unsupervised** learning problem that involves finding groups in data.
- **Density Estimation**: Unsupervised learning problem that involves summarizing the distribution of data.

An example of a clustering algorithm is k-Means where *k* refers to the number of clusters to discover in the data. An example of a density estimation algorithm is Kernel Density Estimation that involves using small groups of closely related data samples to estimate the distribution for new points in the problem space.

> *The most common unsupervised learning task is clustering: detecting potentially useful clusters of input examples. For example, a taxi agent might gradually develop a concept of “good traffic days” and “bad traffic days” without ever being given labeled examples of each by a teacher.*

— Pages 694-695, Artificial Intelligence: A Modern Approach, 3rd edition, 2015.

Clustering and density estimation may be performed to learn about the patterns in the data.

Additional unsupervised methods may also be used, such as visualization that involves graphing or plotting data in different ways and projection methods that involves reducing the dimensionality of the data.

- **Visualization**: Unsupervised learning problem that involves creating plots of data.
- **Projection**: Unsupervised learning problem that involves creating lower-dimensional representations of data.

An example of a visualization technique would be a scatter plot matrix that creates one scatter plot of each pair of variables in the dataset. An example of a projection method would be Principal Component Analysis that involves summarizing a dataset in terms of eigenvalues and eigenvectors, with linear dependencies removed.

> *The goal in such unsupervised learning problems may be to discover groups of similar examples within the data, where it is called clustering, or to determine the distribution of data within the input space, known as density estimation, or to project the data from a high-dimensional space down to two or three dimensions for the purpose of visualization.*

— Page 3, Pattern Recognition and Machine Learning, 2006.

## 3. Reinforcement Learning

Reinforcement learning describes a class of problems where an agent operates in an environment and must *learn* to operate using feedback.

> *Reinforcement learning is learning what to do — how to map situations to actions—so as to maximize a numerical reward signal. The learner is not told which actions to take, but instead must discover which actions yield the most reward by trying them.*

— Page 1, Reinforcement Learning: An Introduction, 2nd edition, 2018.

The use of an environment means that there is no fixed training dataset, rather a goal or set of goals that an agent is required to achieve, actions they may perform, and feedback about performance toward the goal.

> *Some machine learning algorithms do not just experience a fixed dataset. For example, reinforcement learning algorithms interact with an environment, so there is a feedback loop between the learning system and its experiences.*

— Page 105, Deep Learning, 2016.

It is similar to supervised learning in that the model has some response from which to learn, although the feedback may be delayed and statistically noisy, making it challenging for the agent or model to connect cause and effect.

An example of a reinforcement problem is playing a game where the agent has the goal of getting a high score and can make moves in the game and received feedback in terms of punishments or rewards.

> *In many complex domains, reinforcement learning is the only feasible way to train a program to perform at high levels. For example, in game playing, it is very hard for a human to provide accurate and consistent evaluations of large numbers of positions, which would be needed to train an evaluation function directly from examples. Instead, the program can be told when it has won or lost, and it can use this information to learn an evaluation function that gives reasonably accurate estimates of the probability of winning from any given position.*

— Page 831, Artificial Intelligence: A Modern Approach, 3rd edition, 2015.

Impressive recent results include the use of reinforcement in Google's AlphaGo in out-performing the world's top Go player.

Some popular examples of reinforcement learning algorithms include Q-learning, temporal-difference learning, and deep reinforcement learning.

# Hybrid Learning Problems

The lines between unsupervised and supervised learning is blurry, and there are many hybrid approaches that draw from each field of study.

In this section, we will take a closer look at some of the more common hybrid fields of study: semi-supervised, self-supervised, and multi-instance learning.

## 4. Semi-Supervised Learning

Semi-supervised learning is supervised learning where the training data contains very few labeled examples and a large number of unlabeled examples.

The goal of a semi-supervised learning model is to make effective use of all of the available data, not just the labelled data like in supervised learning.

> *In semi-supervised learning we are given a few labeled examples and must make what we can of a large collection of unlabeled examples. Even the labels themselves may not be the oracular truths that we hope for.*

— Page 695, Artificial Intelligence: A Modern Approach, 3rd edition, 2015.

Making effective use of unlabelled data may require the use of or inspiration from unsupervised methods such as clustering and density estimation. Once groups or patterns are discovered, supervised methods or ideas from supervised learning may be used to label the unlabeled examples or apply labels to unlabeled representations later used for prediction.

> *Unsupervised learning can provide useful cues for how to group examples in representation space. Examples that cluster tightly in the input space should be mapped to similar representations.*

— Page 243, Deep Learning, 2016.

It is common for many real-world supervised learning problems to be examples of semi-supervised learning problems given the expense or computational cost for labeling examples. For example, classifying photographs requires a dataset of photographs that have already been labeled by human operators.

Many problems from the fields of computer vision (image data), natural language processing (text data), and automatic speech recognition (audio data) fall into this category and cannot be easily addressed using standard supervised learning methods.

> *… in many practical applications labeled data is very scarce but unlabeled data is plentiful. "Semisupervised" learning attempts to improve the accuracy of supervised learning by exploiting information in unlabeled data. This sounds like magic, but it can work!*

— Page 467, Data Mining: Practical Machine Learning Tools and Techniques, 4th edition, 2016.

## 5. Self-Supervised Learning

Self-supervised learning refers to an unsupervised learning problem that is framed as a supervised learning problem in order to apply supervised learning algorithms to solve it.

Supervised learning algorithms are used to solve an alternate or pretext task, the result of which is a model or representation that can be used in the solution of the original (actual) modeling problem.

> *The self-supervised learning framework requires only unlabeled data in order to formulate a pretext learning task such as predicting context or image rotation, for which a target objective can be computed without supervision.*

— Revisiting Self-Supervised Visual Representation Learning, 2019.

A common example of self-supervised learning is computer vision where a corpus of unlabeled images is available and can be used to train a supervised model, such as making images grayscale and having a model predict a color representation (colorization) or removing blocks of the image and have a model predict the missing parts (inpainting).

> *In discriminative self-supervised learning, which is the main focus of this work, a model is trained on an auxiliary or 'pretext' task for which ground-truth is available for free. In most cases, the pretext task involves predicting some hidden portion of the data (for example, predicting color for gray-scale images*

— Scaling and Benchmarking Self-Supervised Visual Representation Learning, 2019.

A general example of self-supervised learning algorithms are autoencoders. These are a type of neural network that is used to create a compact or compressed representation of an input sample. They achieve this via a model that has an encoder and a decoder element separated by a bottleneck that represents the internal compact representation of the input.

> *An autoencoder is a neural network that is trained to attempt to copy its input to its output. Internally, it has a hidden layer h that describes a code used to represent the input.*

— Page 502, Deep Learning, 2016.

These autoencoder models are trained by providing the input to the model as both input and the target output, requiring that the model reproduce the input by first encoding it to a compressed representation then decoding it back to the original. Once trained, the decoder is discarded and the encoder is used as needed to create compact representations of input.

Although autoencoders are trained using a supervised learning method, they solve an unsupervised learning problem, namely, they are a type of projection method for reducing the dimensionality of input data.

> *Traditionally, autoencoders were used for dimensionality reduction or feature learning.*

— Page 502, Deep Learning, 2016.

Another example of self-supervised learning is generative adversarial networks, or GANs. These are generative models that are most commonly used for creating synthetic photographs using only a collection of unlabeled examples from the target domain.

GAN models are trained indirectly via a separate discriminator model that classifies examples of photos from the domain as real or fake (generated), the result of which is fed back to update the GAN model and encourage it to generate more realistic photos on the next iteration.

> *The generator network directly produces samples […]. Its adversary, the discriminator network, attempts to distinguish between samples drawn from the training data and samples drawn from the generator. The discriminator emits a probability value given by d(x; θ(d)), indicating the probability that x is a real training example rather than a fake sample drawn from the model.*

— Page 699, Deep Learning, 2016.

## 6. Multi-Instance Learning

Multi-instance learning is a supervised learning problem where individual examples are unlabeled; instead, bags or groups of samples are labeled.

> *In multi-instance learning, an entire collection of examples is labeled as containing or not containing an example of a class, but the individual members of the collection are not labeled.*

— Page 106, Deep Learning, 2016.

Instances are in “*bags*” rather than sets because a given instance may be present one or more times, e.g. duplicates.

Modeling involves using knowledge that one or some of the instances in a bag are associated with a target label, and to predict the label for new bags in the future given their composition of multiple unlabeled examples.

> *In supervised multi-instance learning, a class label is associated with each bag, and the goal of learning is to determine how the class can be inferred from the instances that make up the bag.*

— Page 156, Data Mining: Practical Machine Learning Tools and Techniques, 4th edition, 2016.

Simple methods, such as assigning class labels to individual instances and using standard supervised learning algorithms, often work as a good first step.

## Statistical Inference

Inference refers to reaching an outcome or decision.

In machine learning, fitting a model and making a prediction are both types of inference.

There are different paradigms for inference that may be used as a framework for understanding how some machine learning algorithms work or how some learning problems may be approached.

Some examples of approaches to learning are inductive, deductive, and transductive learning and inference.

### 7. Inductive Learning

Inductive learning involves using evidence to determine the outcome.

Inductive reasoning refers to using specific cases to determine general outcomes, e.g. specific to general.

Most machine learning models learn using a type of inductive inference or inductive reasoning where general rules (the model) are learned from specific historical examples (the data).

> *… the problem of induction, which is the problem of how to draw general conclusions about the future from specific observations from the past.*

— Page 77, Machine Learning: A Probabilistic Perspective, 2012.

Fitting a machine learning model is a process of induction. The model is a generalization of the specific examples in the training dataset.

A model or hypothesis is made about the problem using the training data, and it is believed to hold over new unseen data later when the model is used.

> *Lacking any further information, our assumption is that the best hypothesis regarding unseen instances is the hypothesis that best fits the observed training data. This is the fundamental assumption of inductive learning …*

— Page 23, Machine Learning, 1997.

## 8. Deductive Inference

Deduction or deductive inference refers to using general rules to determine specific outcomes.

We can better understand induction by contrasting it with deduction.

Deduction is the reverse of induction. If induction is going from the specific to the general, deduction is going from the general to the specific.

> *… the simple observation that induction is just the inverse of deduction!*

— Page 291, Machine Learning, 1997.

Deduction is a top-down type of reasoning that seeks for all premises to be met before determining the conclusion, whereas induction is a bottom-up type of reasoning that uses available data as evidence for an outcome.

In the context of machine learning, once we use induction to fit a model on a training dataset, the model can be used to make predictions. The use of the model is a type of deduction or deductive inference.

## 9. Transductive Learning

Transduction or transductive learning is used in the field of statistical learning theory to refer to predicting specific examples given specific examples from a domain.

It is different from induction that involves learning general rules from specific examples, e.g. specific to specific.

> *Induction, deriving the function from the given data. Deduction, deriving the values of the given function for points of interest. Transduction, deriving the values of the unknown function for points of interest from the given data.*

— Page 169, The Nature of Statistical Learning Theory, 1995.

Unlike induction, no generalization is required; instead, specific examples are used directly. This may, in fact, be a simpler problem than induction to solve.

> *The model of estimating the value of a function at a given point of interest describes a new concept of inference: moving from the particular to the particular. We call this type of inference transductive inference. Note that this concept of inference appears when one would like to get the best result from a restricted amount of information.*

— Page 169, The Nature of Statistical Learning Theory, 1995.

A classical example of a transductive algorithm is the k-Nearest Neighbors algorithm that does not model the training data, but instead uses it directly each time a prediction is required.

For more on the topic of transduction, see the tutorial:

- Gentle Introduction to Transduction in Machine Learning

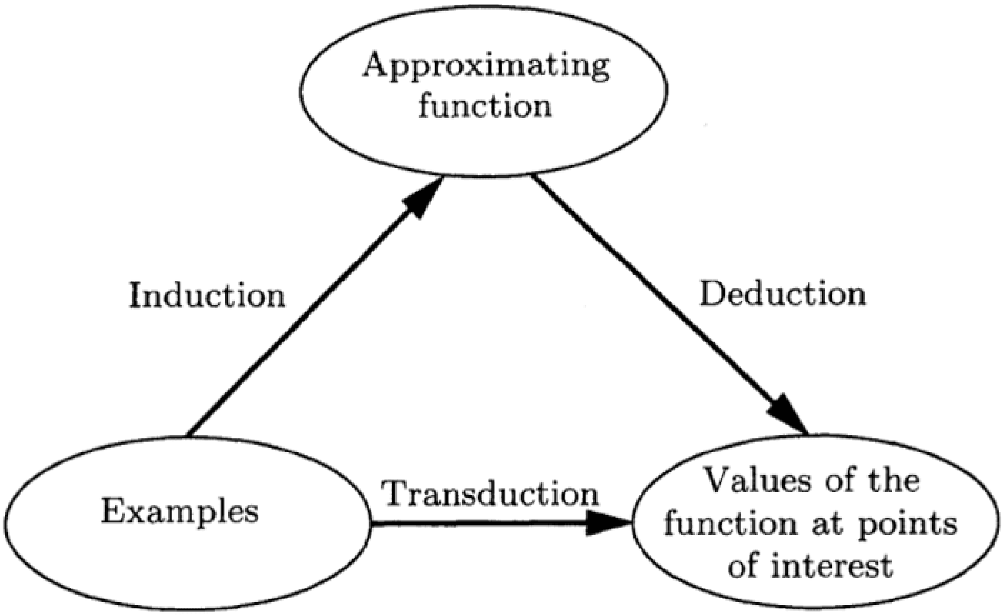
## Contrasting Induction, Deduction, and Transduction:

We can contrast these three types of inference in the context of machine learning.

For example:

- **Induction**: Learning a general model from specific examples.
- **Deduction**: Using a model to make predictions.
- **Transduction**: Using specific examples to make predictions.

The image below summarizes these three different approaches nicely.



Relationship Between Induction, Deduction, and Transduction
Taken from The Nature of Statistical Learning Theory.

## Learning Techniques

There are many techniques that are described as types of learning.

In this section, we will take a closer look at some of the more common methods.

This includes multi-task, active, online, transfer, and ensemble learning.

### 10. Multi-Task Learning

Multi-task learning is a type of supervised learning that involves fitting a model on one dataset that addresses multiple related problems.

It involves devising a model that can be trained on multiple related tasks in such a way that the performance of the model is improved by training across the tasks as compared to being trained on any single task.

> *Multi-task learning is a way to improve generalization by pooling the examples (which can be seen as soft constraints imposed on the parameters) arising out of several tasks.*

— Page 244, Deep Learning, 2016.

Multi-task learning can be a useful approach to problem-solving when there is an abundance of input data labeled for one task that can be shared with another task with much less labeled data.

> *… we may want to learn multiple related models at the same time, which is known as multi-task learning. This will allow us to "borrow statistical strength" from tasks with lots of data and to share it with tasks with little data.*

Page 231, Machine Learning: A Probabilistic Perspective, 2012.

For example, it is common for a multi-task learning problem to involve the same input patterns that may be used for multiple different outputs or supervised learning problems. In this setup, each output may be predicted by a different part of the model, allowing the core of the model to generalize across each task for the same inputs.

> *In the same way that additional training examples put more pressure on the parameters of the model towards values that generalize well, when part of a model is shared across tasks, that part of the model is more constrained towards good values (assuming the sharing is justified), often yielding better generalization.*

— Page 244, Deep Learning, 2016.

A popular example of multi-task learning is where the same word embedding is used to learn a distributed representation of words in text that is then shared across multiple different natural language processing supervised learning tasks.

## 11. Active Learning

Active learning is a technique where the model is able to query a human user operator during the learning process in order to resolve ambiguity during the learning process.

> *Active learning: The learner adaptively or interactively collects training examples, typically by querying an oracle to request labels for new points.*

— Page 7, Foundations of Machine Learning, 2nd edition, 2018.

Active learning is a type of supervised learning and seeks to achieve the same or better performance of so-called "*passive*" supervised learning, although by being more efficient about what data is collected or used by the model.

> *The key idea behind active learning is that a machine learning algorithm can achieve greater accuracy with fewer training labels if it is allowed to choose the data from which it learns. An*

> *active learner may pose queries, usually in the form of unlabeled data instances to be labeled by an oracle (e.g., a human annotator).*

— Active Learning Literature Survey, 2009.

It is not unreasonable to view active learning as an approach to solving semi-supervised learning problems, or an alternative paradigm for the same types of problems.

> *… we see that active learning and semi-supervised learning attack the same problem from opposite directions. While semi-supervised methods exploit what the learner thinks it knows about the unlabeled data, active methods attempt to explore the unknown aspects. It is therefore natural to think about combining the two*

— Active Learning Literature Survey, 2009.

Active learning is a useful approach when there is not much data available and new data is expensive to collect or label.

The active learning process allows the sampling of the domain to be directed in a way that minimizes the number of samples and maximizes the effectiveness of the model.

> *Active learning is often used in applications where labels are expensive to obtain, for example computational biology applications.*

— Page 7, Foundations of Machine Learning, 2nd edition, 2018.

## 12. Online Learning

Online learning involves using the data available and updating the model directly before a prediction is required or after the last observation was made.

Online learning is appropriate for those problems where observations are provided over time and where the probability distribution of observations is expected to also change over time. Therefore, the model is expected to change just as frequently in order to capture and harness those changes.

> *Traditionally machine learning is performed offline, which means we have a batch of data, and we optimize an equation […] However, if we have streaming data, we need to perform online learning, so we can update our estimates as each new data point arrives rather than waiting until "the end" (which may never occur).*

— Page 261, Machine Learning: A Probabilistic Perspective, 2012.

This approach is also used by algorithms where there may be more observations than can reasonably fit into memory, therefore, learning is performed incrementally over observations, such as a stream of data.

> *Online learning is helpful when the data may be changing rapidly over time. It is also useful for applications that involve a large collection of data that is constantly growing, even if changes are gradual.*

— Page 753, Artificial Intelligence: A Modern Approach, 3rd edition, 2015.

Generally, online learning seeks to minimize "*regret*," which is how well the model performed compared to how well it might have performed if all the available information was available as a batch.

> *In the theoretical machine learning community, the objective used in online learning is the regret, which is the averaged loss incurred relative to the best we could have gotten in hindsight using a single fixed parameter value*

— Page 262, Machine Learning: A Probabilistic Perspective, 2012.

One example of online learning is so-called stochastic or online gradient descent used to fit an artificial neural network.

> *The fact that stochastic gradient descent minimizes generalization error is easiest to see in the online learning case, where examples or minibatches are drawn from a stream of data.*

— Page 281, Deep Learning, 2016.

## 13. Transfer Learning

Transfer learning is a type of learning where a model is first trained on one task, then some or all of the model is used as the starting point for a related task.

> *In transfer learning, the learner must perform two or more different tasks, but we assume that many of the factors that explain the variations in P1 are relevant to the variations that need to be captured for learning P2.*

— Page 536, Deep Learning, 2016.

It is a useful approach on problems where there is a task related to the main task of interest and the related task has a large amount of data.

It is different from multi-task learning as the tasks are learned sequentially in transfer learning, whereas multi-task learning seeks good performance on all considered tasks by a single model at the same time in parallel.

> *… pretrain a deep convolutional net with 8 layers of weights on a set of tasks (a subset of the 1000 ImageNet object categories) and then initialize a same-size network with the first k layers of the first net. All the layers of the second network (with the upper layers initialized randomly) are then jointly trained to perform a different set of tasks (another subset of the 1000 ImageNet object categories), with fewer training examples than for the first set of tasks.*

— Page 325, Deep Learning, 2016.

An example is image classification, where a predictive model, such as an artificial neural network, can be trained on a large corpus of general images, and the weights of the model can be used as a starting point when training on a smaller more specific dataset, such as dogs and cats. The features already learned by the model on the broader task, such as extracting lines and patterns, will be helpful on the new related task.

> *If there is significantly more data in the first setting (sampled from P1), then that may help to learn representations that are useful to quickly generalize from only very few examples drawn from P2. Many visual categories share low-level notions of edges and visual shapes, the effects of geometric changes, changes in lighting, etc.*

— Page 536, Deep Learning, 2016.

As noted, transfer learning is particularly useful with models that are incrementally trained and an existing model can be used as a starting point for continued training, such as deep learning networks.

For more on the topic of transfer learning, see the tutorial:

- A Gentle Introduction to Transfer Learning for Deep Learning

## 14. Ensemble Learning

Ensemble learning is an approach where two or more modes are fit on the same data and the predictions from each model are combined.

> *The field of ensemble learning provides many ways of combining the ensemble members' predictions, including uniform weighting and weights chosen on a validation set.*

— Page 472, Deep Learning, 2016.

The objective of ensemble learning is to achieve better performance with the ensemble of models as compared to any individual model. This involves both deciding how to create models used in the ensemble and how to best combine the predictions from the ensemble members.

> *Ensemble learning can be broken down into two tasks: developing a population of base learners from the training data, and then combining them to form the composite predictor.*

— Page 605, The Elements of Statistical Learning: Data Mining, Inference, and Prediction, 2nd edition, 2016.

Ensemble learning is a useful approach for improving the predictive skill on a problem domain and to reduce the variance of stochastic learning algorithms, such as artificial neural networks.

Some examples of popular ensemble learning algorithms include: weighted average, stacked generalization (stacking), and bootstrap aggregation (bagging).

> *Bagging, boosting, and stacking have been developed over the last couple of decades, and their performance is often astonishingly good. Machine learning researchers have struggled to understand why.*

— Page 480, Data Mining: Practical Machine Learning Tools and Techniques, 4th edition, 2016.

For more on the topic of ensemble learning, see the tutorial:

- Ensemble Learning Methods for Deep Learning Neural Networks

# Further Reading

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

## Books

- Pattern Recognition and Machine Learning, 2006.
- Deep Learning, 2016.
- Reinforcement Learning: An Introduction, 2nd edition, 2018.
- Data Mining: Practical Machine Learning Tools and Techniques, 4th edition, 2016.
- The Elements of Statistical Learning: Data Mining, Inference, and Prediction, 2nd edition, 2016.
- Machine Learning: A Probabilistic Perspective, 2012.
- Machine Learning, 1997.
- The Nature of Statistical Learning Theory, 1995.
- Foundations of Machine Learning, 2nd edition, 2018.
- Artificial Intelligence: A Modern Approach, 3rd edition, 2015.

## Papers

- Revisiting Self-Supervised Visual Representation Learning, 2019.
- Active Learning Literature Survey, 2009.