Machine Learning is the science (and art) of programming computers so they can learn from data.

Here is a slightly more general definition:
Machine Learning is the field of study that gives computers the ability to learn without being explicitly programmed.
—Arthur Samuel, *1959*

And a more engineering-oriented one:
A computer program is said to learn from experience E with respect to some task T and some performance measure P, if its performance on T, as measured by P, improves with experience E.
—Tom Mitchell, *1997*

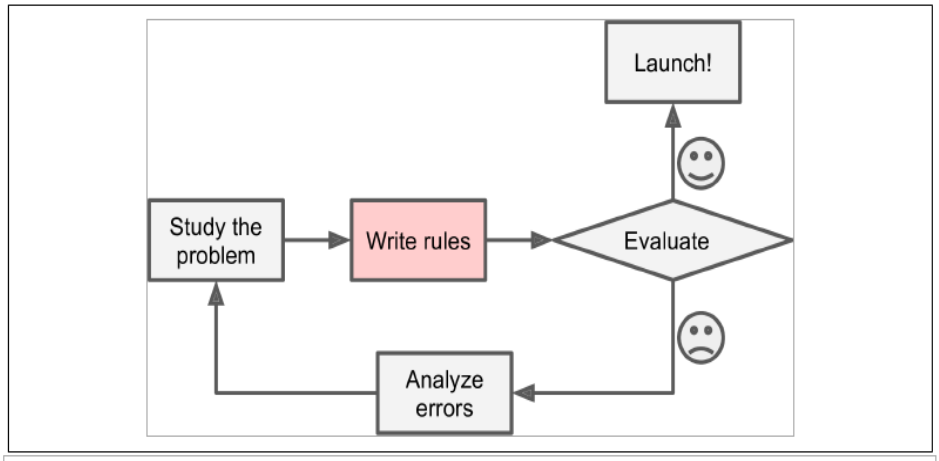


Figure 1-1. The traditional approach

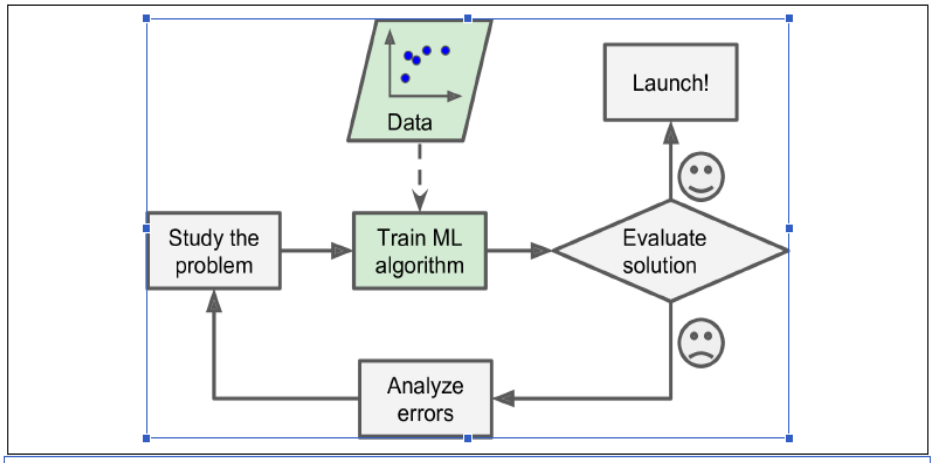


Figure 1-2. Machine Learning approach

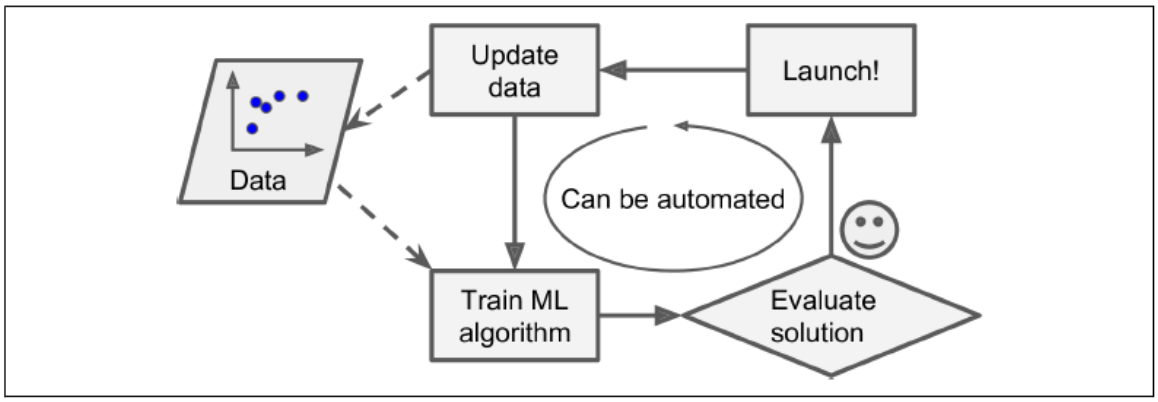


Figure 1-3. Automatically adapting to change

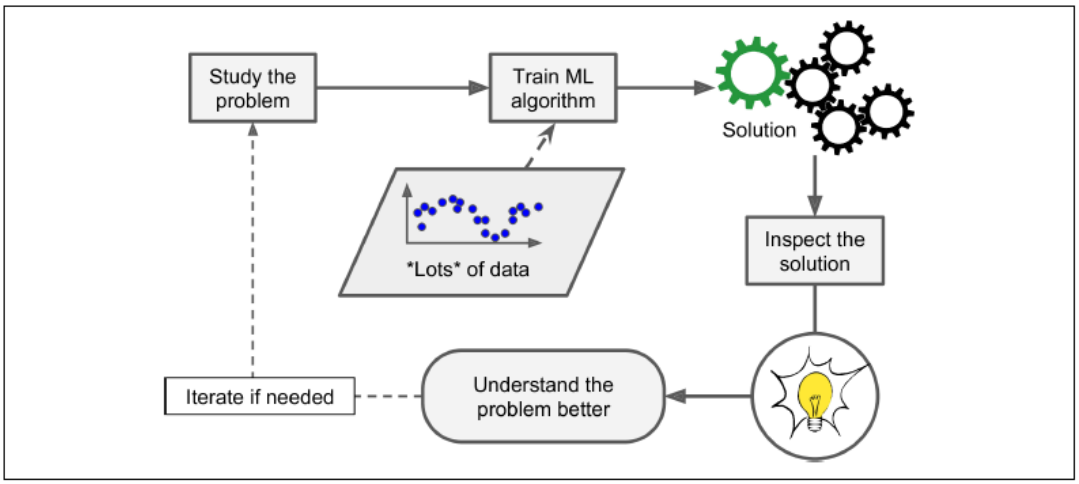


Figure 1-4. Machine Learning can help humans learn

ML Types/classification

• Whether or not they are trained with human supervision (supervised, unsupervised, semisupervised, and Reinforcement Learning)

• Whether or not they can learn incrementally on the fly (online versus batch learning)

• Whether they work by simply comparing new data points to known data points, or instead detect patterns in the training data and build a predictive model, much like scientists do (instance-based versus model-based learning)

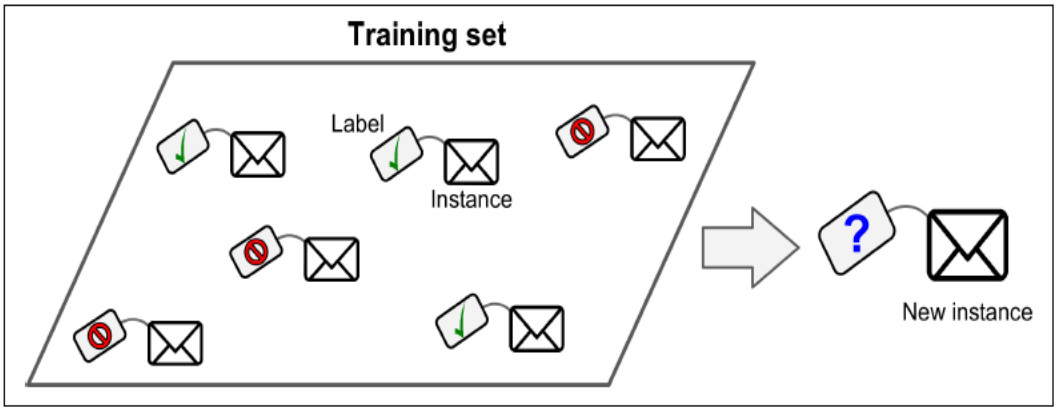
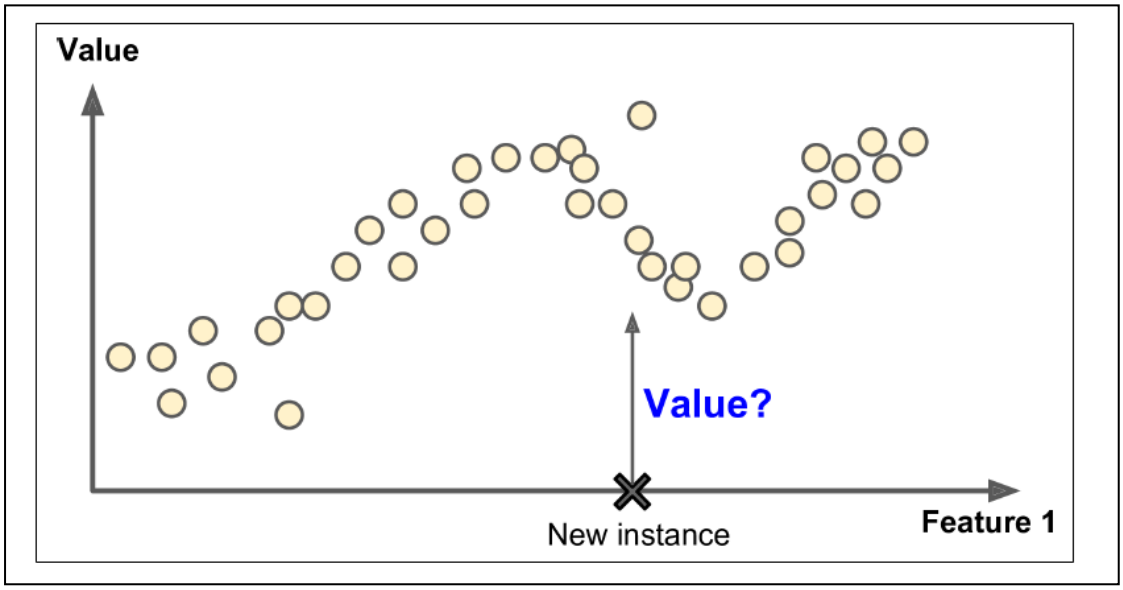


Figure 1-5. A labeled training set for supervised learning (e.g., spam classification)


Value
Value?
New instance
Feature 1

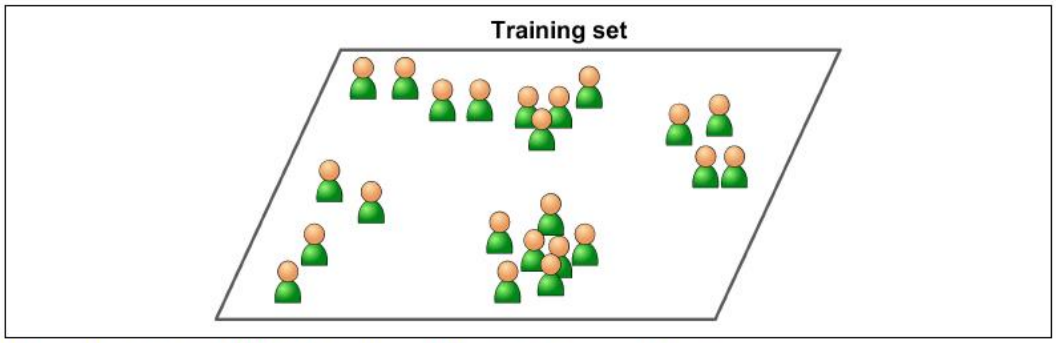


Figure 1-7. An unlabeled training set for unsupervised learning

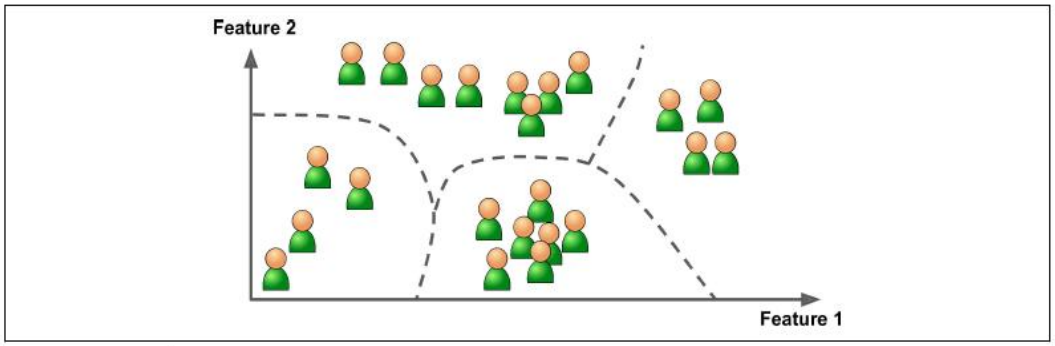


Figure 1-8. Clustering

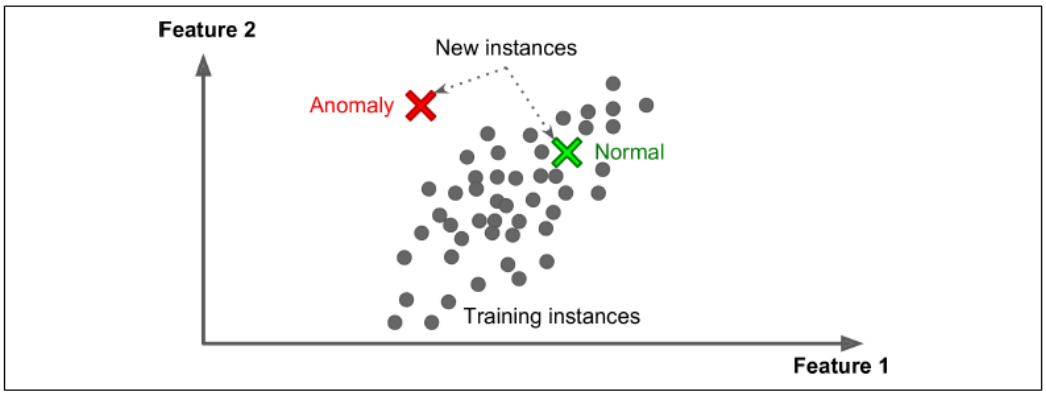


Figure 1-10. Anomaly detection

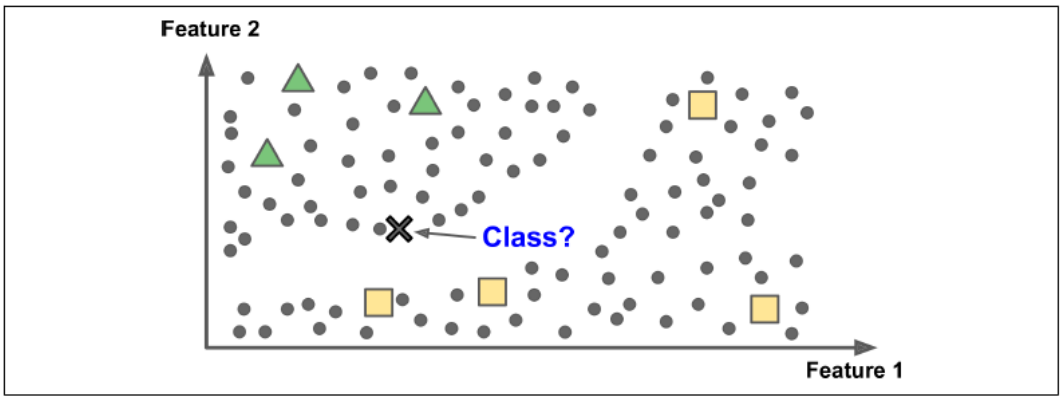
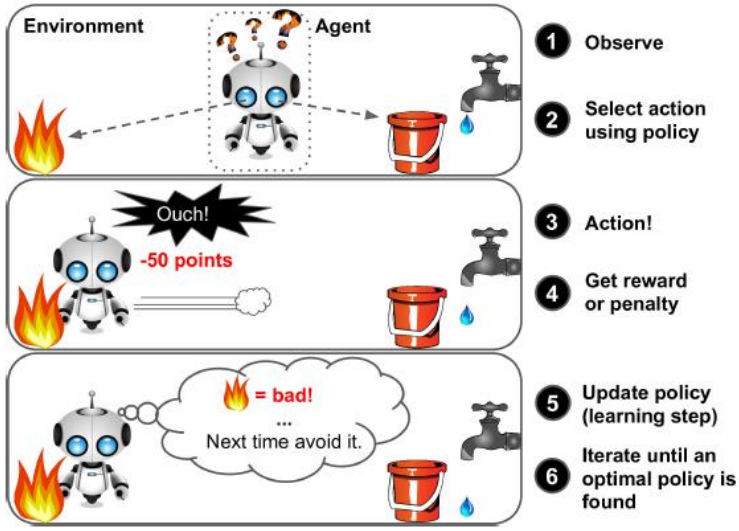


Figure 1-11. Semisupervised learning


Environment
Agent
Ouch!
-50 points
= bad!
...
Next time avoid it.
1 Observe
2 Select action using policy
3 Action!
4 Get reward or penalty
5 Update policy (learning step)
6 Iterate until an optimal policy is found

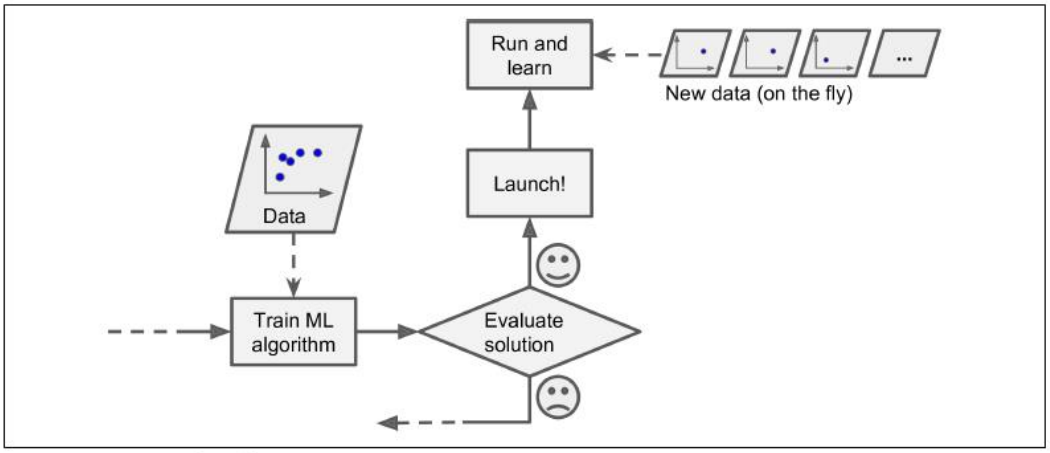


Figure 1-13. Online learning

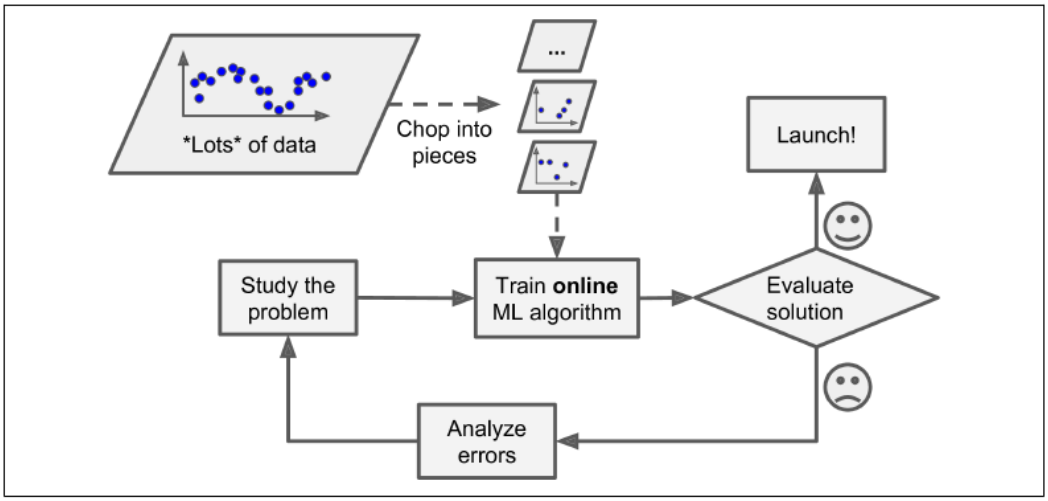


Figure 1-14. Using online learning to handle huge datasets

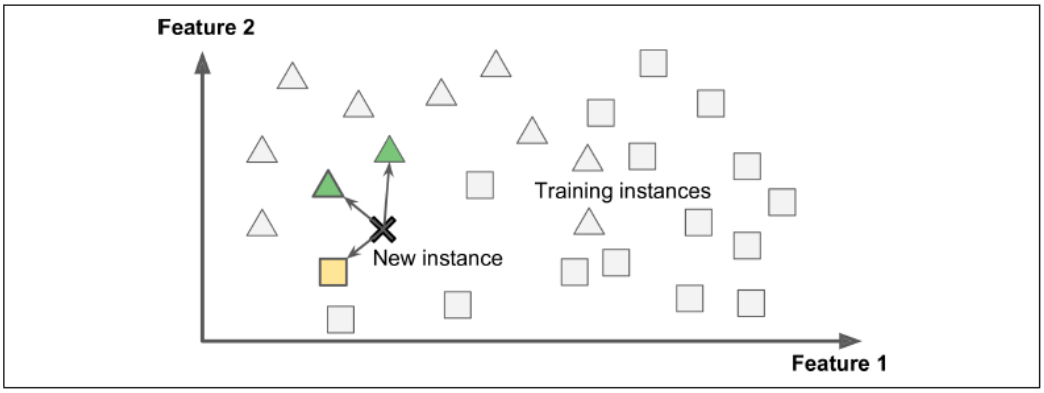


Figure 1-15. Instance-based learning

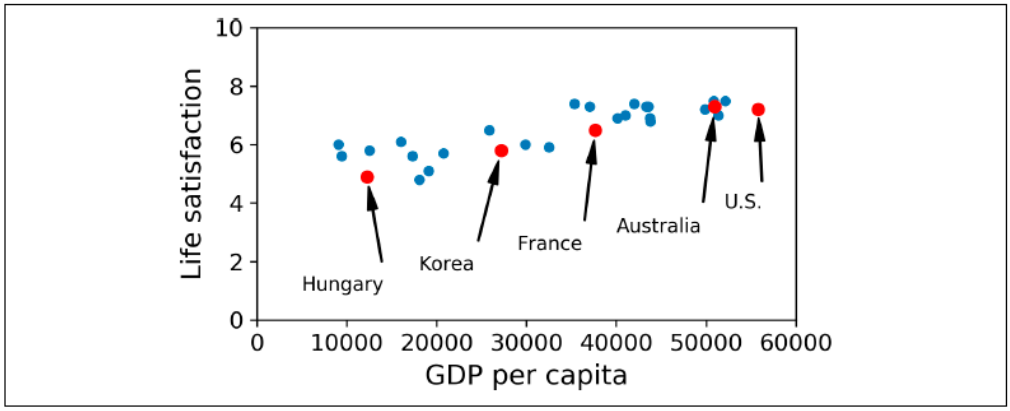


Figure 1-17. Do you see a trend here?

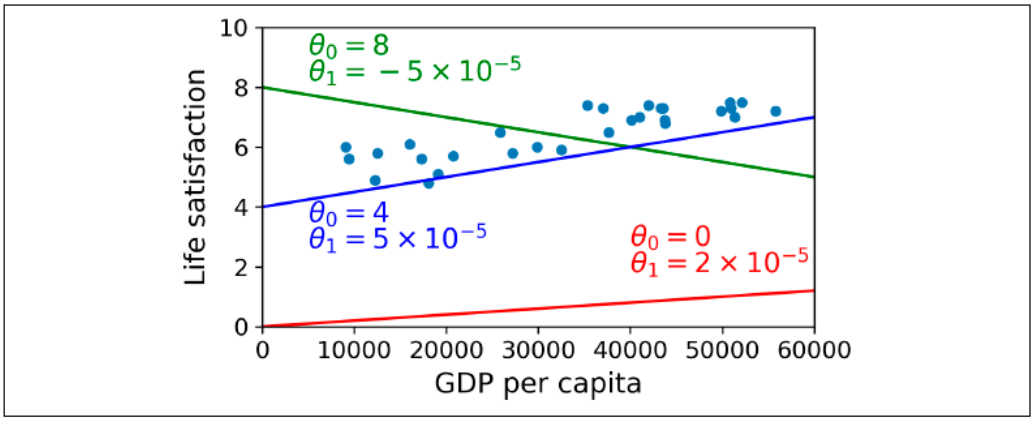


Figure 1-18. A few possible linear models

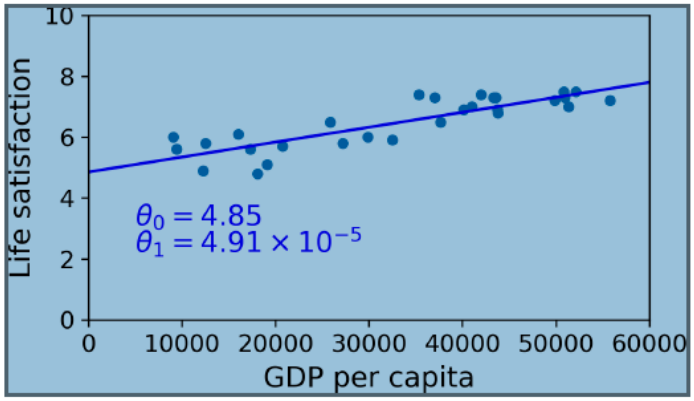


Figure 1-19. The linear model that fits the training data best

The Unreasonable Effectiveness of Data

In a famous paper published in 2001, Microsoft researchers Michele Banko and Eric Brill showed that very different Machine Learning algorithms, including fairly simple ones, performed almost identically well on a complex problem of natural language disambiguation[8] once they were given enough data (as you can see in Figure 1-20).

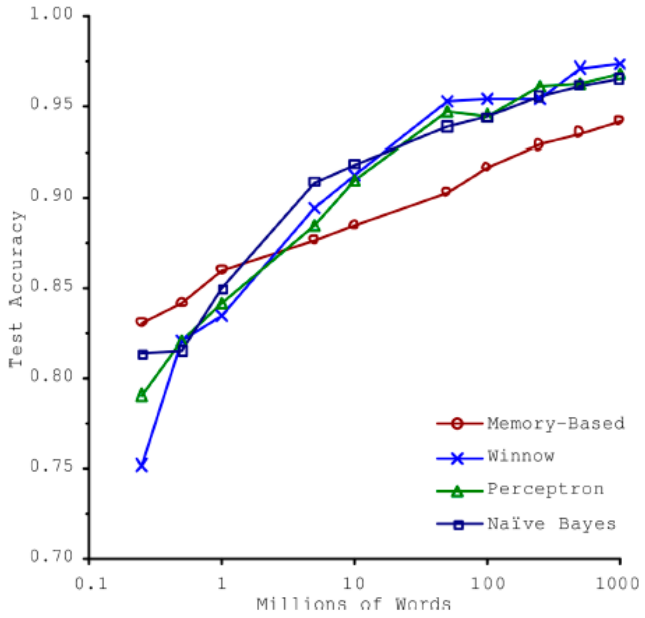


Figure 1-20. The importance of data versus algorithms[9]

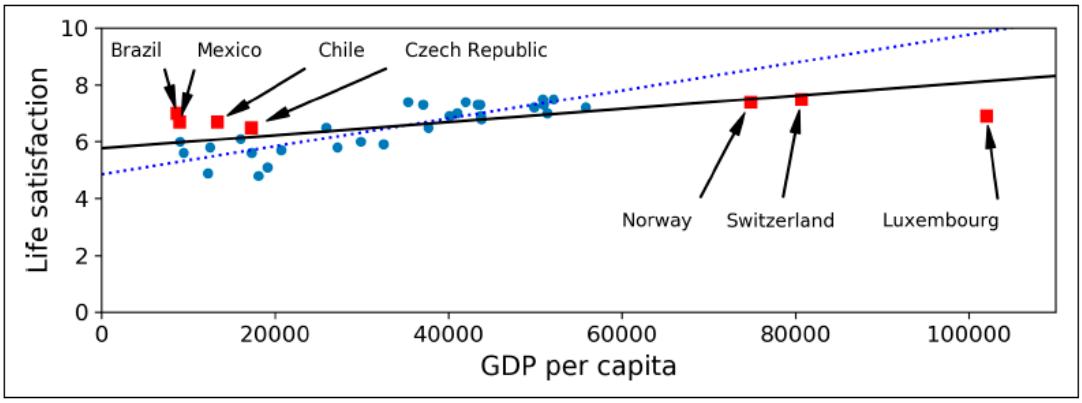


Figure 1-21. A more representative training sample

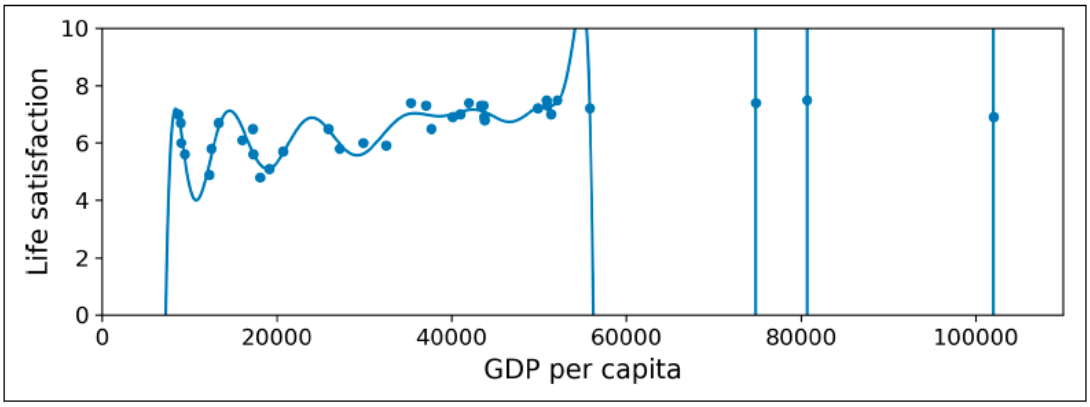


Figure 1-22. Overfitting the training data

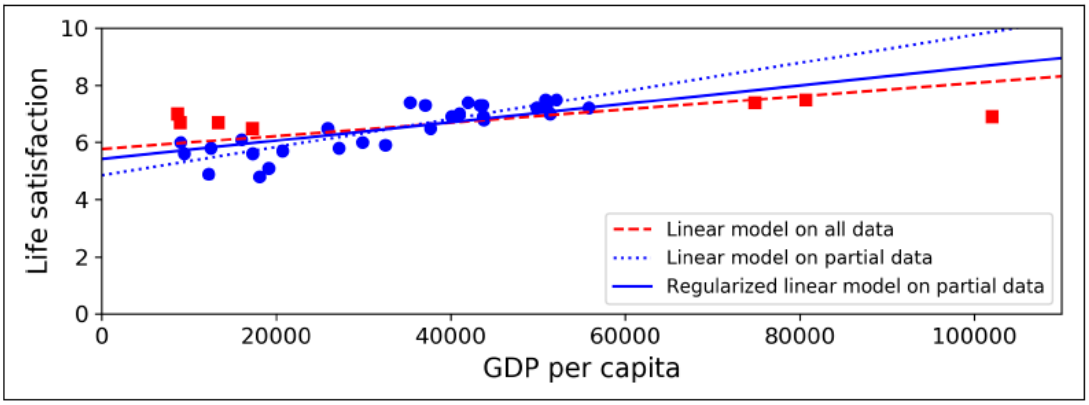


Figure 1-23. Regularization reduces the risk of overfitting