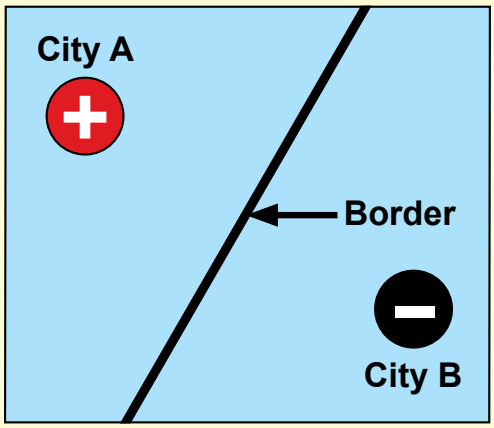
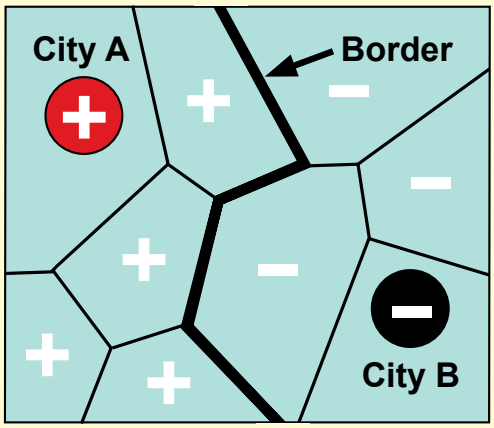
Analogy Algorithms

The analogist approach learns by finding an example that is sufficiently similar to the case in the data under consideration. At the most primitive level it uses the *nearest-neighbor* algorithm, also referred to as the *lazy learner*. For example, when a Facebook user uploads a photograph of a person how does Facebook recognize the image as being the face of a person? It looks through its database of images to find another image that contains enough similar features to make it highly probable that it is of the same kind. If this image is known by Facebook to be the face of a person, then the just uploaded photograph is also classified as containing the face of a person.

Pedro Domingos in his book *The Master Algorithm* (Basic Books 2015, page 180-181) provides the following example of how the apparently quite primitive capabilities of a *lazy learner* can be applied in a surprisingly sophisticated manner. Let us assume that we want to determine the approximate border between two states in some country. The lazy learner might start with the hypothesis that the border is a straight-line half-way between the capital cities of the two states. By taking into consideration a large number of towns on either side of the border it is then able to construct an intricate border based on just the location of each town and the state that the town belongs to.





K-nearest-neighbor is a refinement of the nearest-neighbor algorithm by taking into account several nearest neighbors instead of just one nearest neighbor. For example, if the first nearest neighbor of the case being analyzed is sufficiently like the test case but the next two nearest neighbors are not then the combined vote would be that the test case classification is not appropriate for the case under consideration.

The two most difficult problems encountered by analogical learning is how to measure similarity and what to infer from the similarity. A typical example is an automated Help Desk. Chances are that the problem that a customer is encountering with a product has occurred previously and that a solution has been successfully implemented.