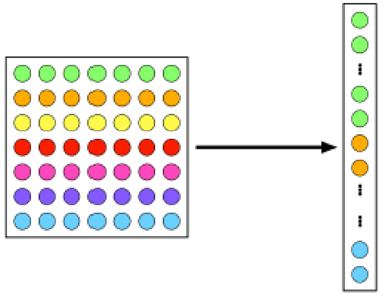
Data Representation as Vectors

- Learning algorithms work with data given as a set of input-output pairs $\{(\boldsymbol{x}_n, y_n)\}_{n=1}^N$ (supervised), or as a set of inputs $\{\boldsymbol{x}_n\}_{n=1}^N$ (unsupervised)
- Each $\boldsymbol{x}_n$ is usually a D dimensional feature vector containing the values of D features/attributes that encode properties of the data it represents, e.g.,
    - Representing a 7×7 image: $\boldsymbol{x}_n$ can be a 49×1 vector of pixel intensities



    - Note: Good features can also be learned from data (feature learning) or extracted using hand-crafted rules defined by a domain expert

Distances:

- Each feature vector $\boldsymbol{x}_n \in \mathbb{R}^{D \times 1}$ is a point in a D dim. vector space
- Standard rules of vector algebra apply on such representations, e.g.,
  - Euclidean distance b/w points $\boldsymbol{x}_n \in \mathbb{R}^D$ and $\boldsymbol{x}_m \in \mathbb{R}^D$

$$d(\boldsymbol{x}_n, \boldsymbol{x}_m) = ||\boldsymbol{x}_n - \boldsymbol{x}_m|| = \sqrt{(\boldsymbol{x}_n - \boldsymbol{x}_m)^\top (\boldsymbol{x}_n - \boldsymbol{x}_m)} = \sqrt{\sum_{d=1}^{D} (x_{nd} - x_{md})^2}$$

  - Inner-product similarity b/w $\boldsymbol{x}_n$ and $\boldsymbol{x}_m$ (cosine, $\boldsymbol{x}_n$, $\boldsymbol{x}_m$ are unit-length vectors)

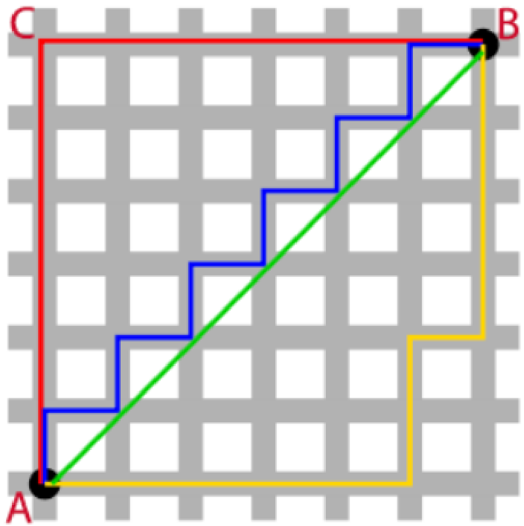
$$s(\boldsymbol{x}_n, \boldsymbol{x}_m) = \boldsymbol{x}_n^\top \boldsymbol{x}_m = \sum_{d=1}^{D} x_{nd} x_{md}$$

  - ℓ_1 distance between two points $\boldsymbol{x}_n$ and $\boldsymbol{x}_m$

$$d_1(\boldsymbol{x}_n, \boldsymbol{x}_m) = ||\boldsymbol{x}_n - \boldsymbol{x}_m||_1 = \sum_{d=1}^{D} |x_{nd} - x_{md}|$$

Manhattan Distance

Travel the green line from A to B in the image to the right and you have traveled the shortest distance from A to B.



If this is part of Manhattan in New York City and you are a crow, you can fly that green path. But if you are an Uber or Lyft driver, you have to keep your car on the street. In that case, you have several choices on how to drive, for example, follow the red line, the blue line, or the yellow line.

The *Manhattan distance* between two points on a grid is:

The sum of the vertical and horizontal distances between them.

Thus, in the image to the right, the Manhattan distance from A to B is the sum of the distance from A to C (upper left corner) and the distance from C to B.

It should be clear that the lengths of the blue and yellow paths are also the Manhattan distance.