ECE 417 Lecture 2: Metric (=Norm) Learning

Mark Hasegawa-Johnson

8/31/2017

Today’s Lecture

- Similarity and Dissimilarity of vectors: all you need is a norm
- Example: the Minkowski Norm (Lp norm)
- Cosine Similarity: you need a dot product
- Example: Diagonal Mahalanobis Distance
- What is Similarity?
- Metric Learning

Norm (or Metric, or Length) of a vector

A norm is:

1. Non-negative, $\|\vec{x}\| \geq 0$
2. Positive definite, $\|\vec{x}\| = 0$ iff $\vec{x} = \vec{0}$
3. Absolute homogeneous, $\|a\vec{x}\| = |a|\|\vec{x}\|$
4. Satisfies the triangle inequality, $\|\vec{x} + \vec{y}\| \leq \|\vec{x}\|+\|\vec{y}\|$

Notice that, from 3 and 4 together, we get

$\|\vec{x} - \vec{y}\| \leq \|\vec{x}\|+\|\vec{y}\|$

Distance between two vectors

The distance between two vectors is just the norm of their difference.

Notice that, because of non-negativity, homogeneity, and triangle inequality, we can write that

$0 \leq ||\vec{x} - \vec{y}|| \leq ||\vec{x}||+||\vec{y}||$

And because of positive definiteness, we also know that $0 = ||\vec{x} - \vec{y}||$ only if $\vec{x} = \vec{y}$.

And the maximum value of $||\vec{x} - \vec{y}||$ is $||\vec{x}||+||\vec{y}||$ achieved only if y is proportional to -x

Today’s Lecture

- Similarity and Dissimilarity of vectors: all you need is a norm
- Example: the Minkowski Norm (Lp norm)
- Cosine Similarity: you need a dot product
- Example: Diagonal Mahalanobis Distance
- What is Similarity?
- Metric Learning

Example: Euclidean (L2) Distance

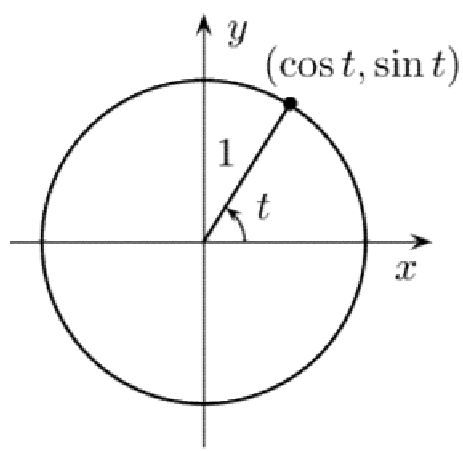
The Euclidean (L2) distance between two vectors is defined as

$$||\vec{x} - \vec{y}||_2 = \sqrt{|x_1 - y_1|^2 + \cdots + |x_D - y_D|^2}$$

1. Non-negative: well, obviously
2. Positive definite: also obvious
3. Absolute homogeneous: easy to show
4. Triangle inequality: easy to show: square both sides of that equation $||\vec{x} - \vec{y}|| \leq ||\vec{x}||+||\vec{y}||$

Example: Euclidean (L2) Distance

Here are the vectors $\vec{x}$, in 2-dimensional space, that have $\|\vec{x}\|_2 = 1$



Attribution: Gustavb, https://commons.wikimedia.org/wiki/File:Unit_circle.svg

Example: Minkowski (Lp) Norm

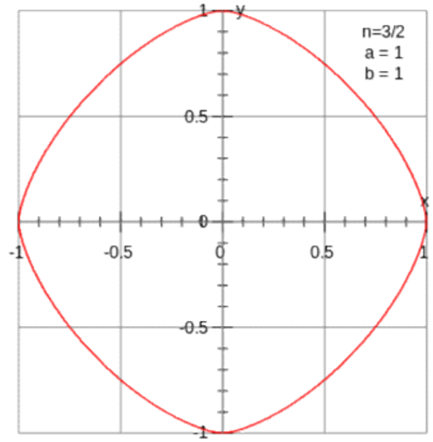
The Minkowski (Lp) distance between two vectors is defined as

$$\|\vec{x}-\vec{y}\|_p = \sqrt[p]{|x_1-y_1|^p + \cdots + |x_D-y_D|^p}$$

1. Non-negative: well, obviously
2. Positive definite: also obvious
3. Absolute homogeneous: easy to show
4. Triangle inequality: easy to show for any particular positive integer value of p (just raise both sides of the equation $\|\vec{x}-\vec{y}\| \leq \|\vec{x}\|$+$\|\vec{y}\|$ to the power of p).

Example: Minkowski (Lp) Distance

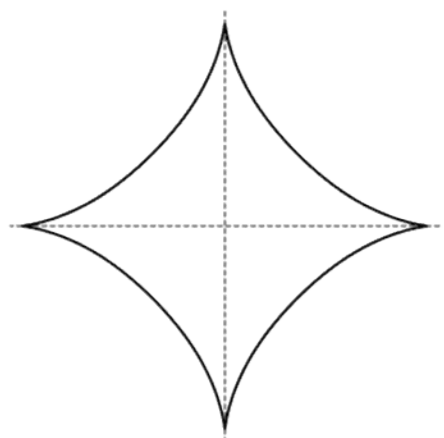
Here are the vectors $\vec{x}$, in 2-dimensional space, that have $\|\vec{x}\|_{3/2} = 1$



Attribution: Krishnavedala,
https://en.wikipedia.org/wiki/Lp_space#/media/File:Superellipse_rounded_diamond.svg

Example: Minkowski (Lp) Distance

Here are the vectors $\vec{x}$, in 2-dimensional space, that have $\|\vec{x}\|_{2/3} = 1$



Attribution: Joelholdsworth, https://commons.wikimedia.org/wiki/File:Astroid.svg

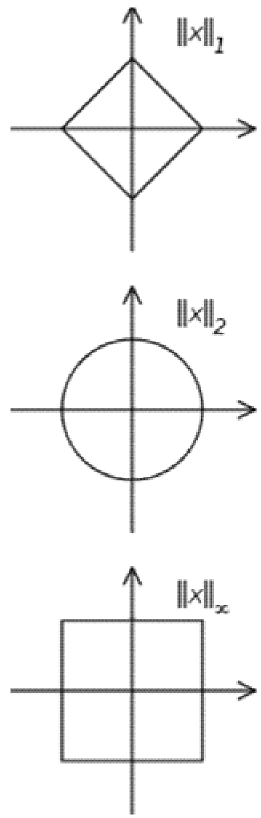
Manhattan Distance and L-infinity Distance

The Manhattan (L1) distance is

$$\|\vec{x}-\vec{y}\|_1 = |x_1 - y_1| + \cdots + |x_D - y_D|$$

The L-infinity distance is

$$\|\vec{x}-\vec{y}\|_\infty = \lim_{p\to\infty} \sqrt[p]{|x_1 - y_1|^p + \cdots + |x_D - y_D|^p}$$
$$= \max_{1\le d\le D} |x_d - y_d|$$



Attribution: Esmil, https://commons.wikimedia.org/wiki/File:Vector_norms.svg

Today’s Lecture

- Similarity and Dissimilarity of vectors: all you need is a norm
- Example: the Minkowski Norm (Lp norm)
- Cosine Similarity: you need a dot product
- Example: Diagonal Mahalanobis Distance
- What is Similarity?
- Metric Learning

Dot product defines a norm

- The dot product between two real-valued vectors is symmetric and linear, so:

$$(\vec{x}-\vec{y})^T(\vec{x}-\vec{y}) = x^T\vec{x} - 2\vec{y}^T\vec{x} + \vec{y}^T\vec{y}$$

(for complex-valued vectors, things are a bit more complicated, but not too much).

- Dot product is always positive definite:

$$(\vec{x}-\vec{y})^T(\vec{x}-\vec{y}) \geq 0$$

$$(\vec{x}-\vec{y})^T(\vec{x}-\vec{y}) = 0 \text{ only if } \vec{x} = \vec{y}$$

- So **a dot product defines a norm**:

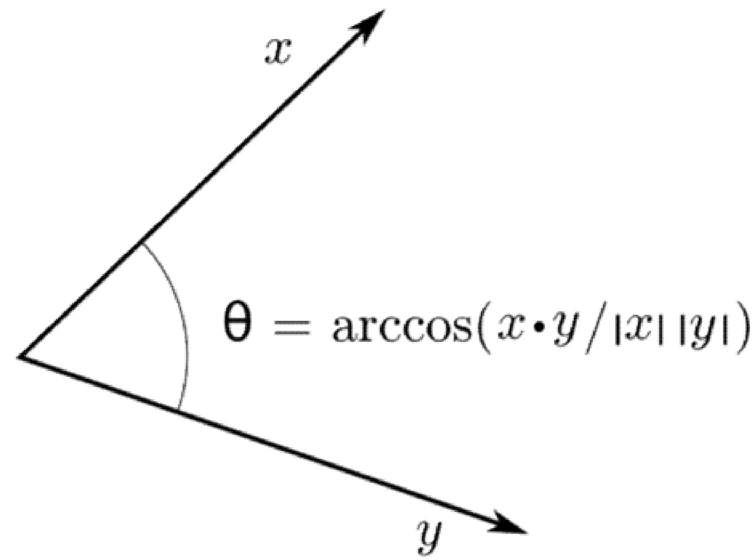
$$\|\vec{x}-\vec{y}\|^2 = (\vec{x}-\vec{y})^T(\vec{x}-\vec{y})$$

$$\|\vec{x}-\vec{y}\|^2 = \|\vec{x}\|^2 - 2\vec{y}^T\vec{x} + \|\vec{y}\|^2$$

Cosine

The cosine of the angle between two vectors is

$$\cos(\vec{x},\vec{y}) = \frac{\vec{y}^T\vec{x}}{\|\vec{x}\|\|\vec{y}\|}$$



Attribution: CSTAR, https://commons.wikimedia.org/wiki/File:Inner-product-angle.png

Today’s Lecture

- Similarity and Dissimilarity of vectors: all you need is a norm
- Example: the Minkowski Norm (Lp norm)
- Cosine Similarity: you need a dot product
- Example: Diagonal Mahalanobis Distance
- What is Similarity?
- Metric Learning

Example: Euclidean distance

- The Euclidean dot product is:

$$\vec{y}^T\vec{x} = x_1y_1 + \cdots + x_Dy_D$$

- The Euclidean distance is:

$$\begin{aligned}\|\vec{x} - \vec{y}\|^2 &= (x_1 - y_1)^2 + \cdots + (x_D - y_D)^2 \\ &= (x_1^2 - 2x_1y_1 + y_1^2) + \cdots + (x_D^2 - 2x_Dy_D + y_D^2) \\ &= \|\vec{x}\|^2 + \|\vec{y}\|^2 - 2(x_1y_1 + \cdots + x_Dy_D)\end{aligned}$$

Example: Mahalanobis Distance

- Suppose that Σ is a diagonal matrix,

$$\Sigma = \begin{bmatrix} \sigma_1^2 & 0 & 0 \\ 0 & \dots & 0 \\ 0 & 0 & \sigma_D^2 \end{bmatrix}, \qquad \Sigma^{-1} = \begin{bmatrix} \frac{1}{\sigma_1^2} & 0 & 0 \\ 0 & \dots & 0 \\ 0 & 0 & \frac{1}{\sigma_D^2} \end{bmatrix}$$

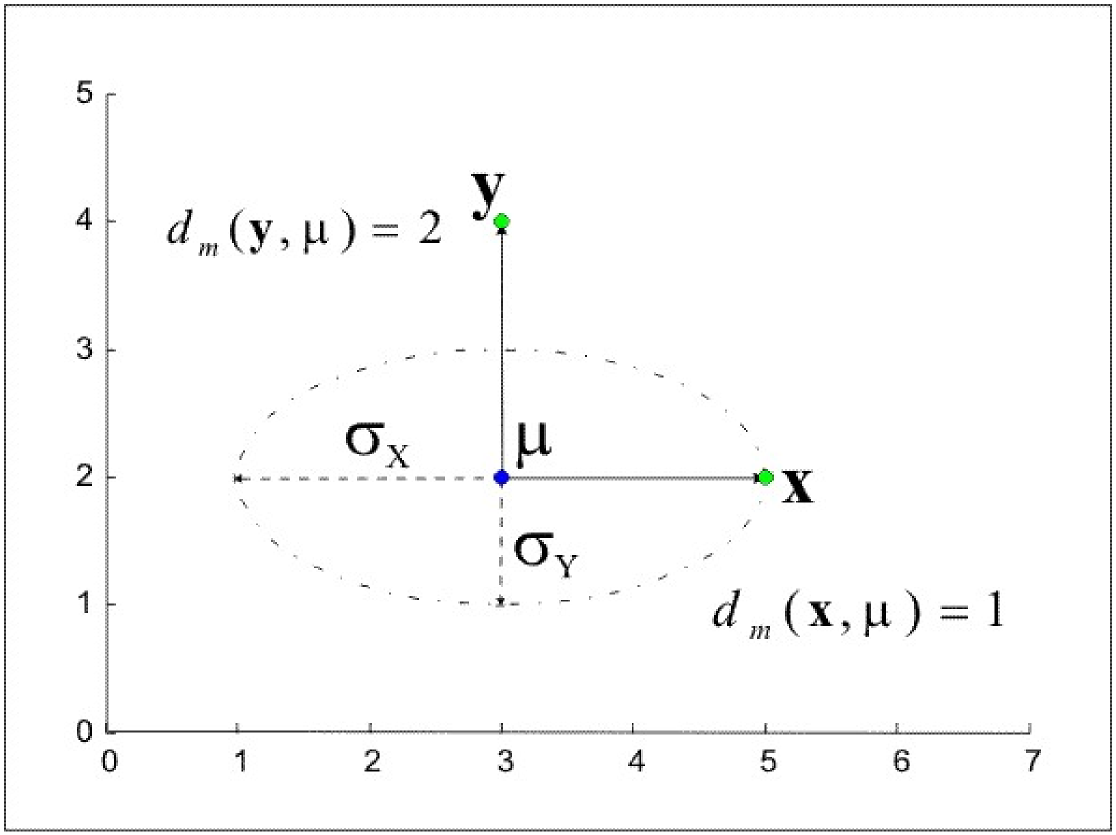
- The Mahalanobis dot product is then defined as:

$$\vec{y}^T \Sigma^{-1} \vec{x} = \frac{x_1 y_1}{\sigma_1^2} + \dots + \frac{x_D y_D}{\sigma_D^2}$$

- The squared Mahalonobis distance is:

$$d_m^2(x, y) = (\vec{x} - \vec{y})^T \Sigma^{-1} (\vec{x} - \vec{y}) = \frac{(x_1 - y_1)^2}{\sigma_1^2} + \dots + \frac{(x_D - y_D)^2}{\sigma_D^2}$$

Example: Mahalanobis Distance

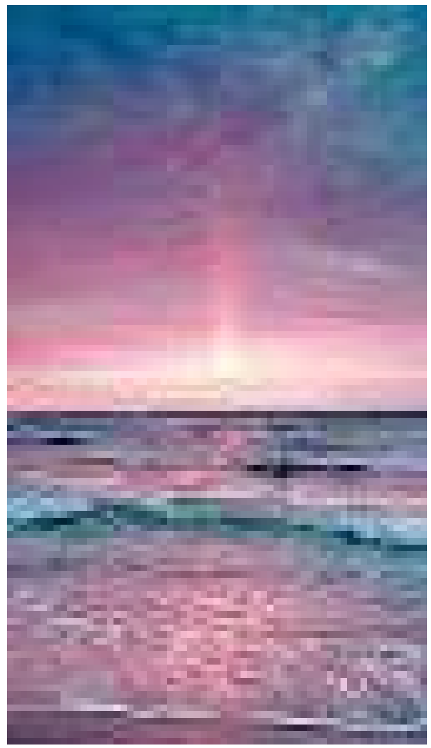


Attribution: Piotrg, https://commons.wikimedia.org/wiki/File:MahalanobisDist1.png

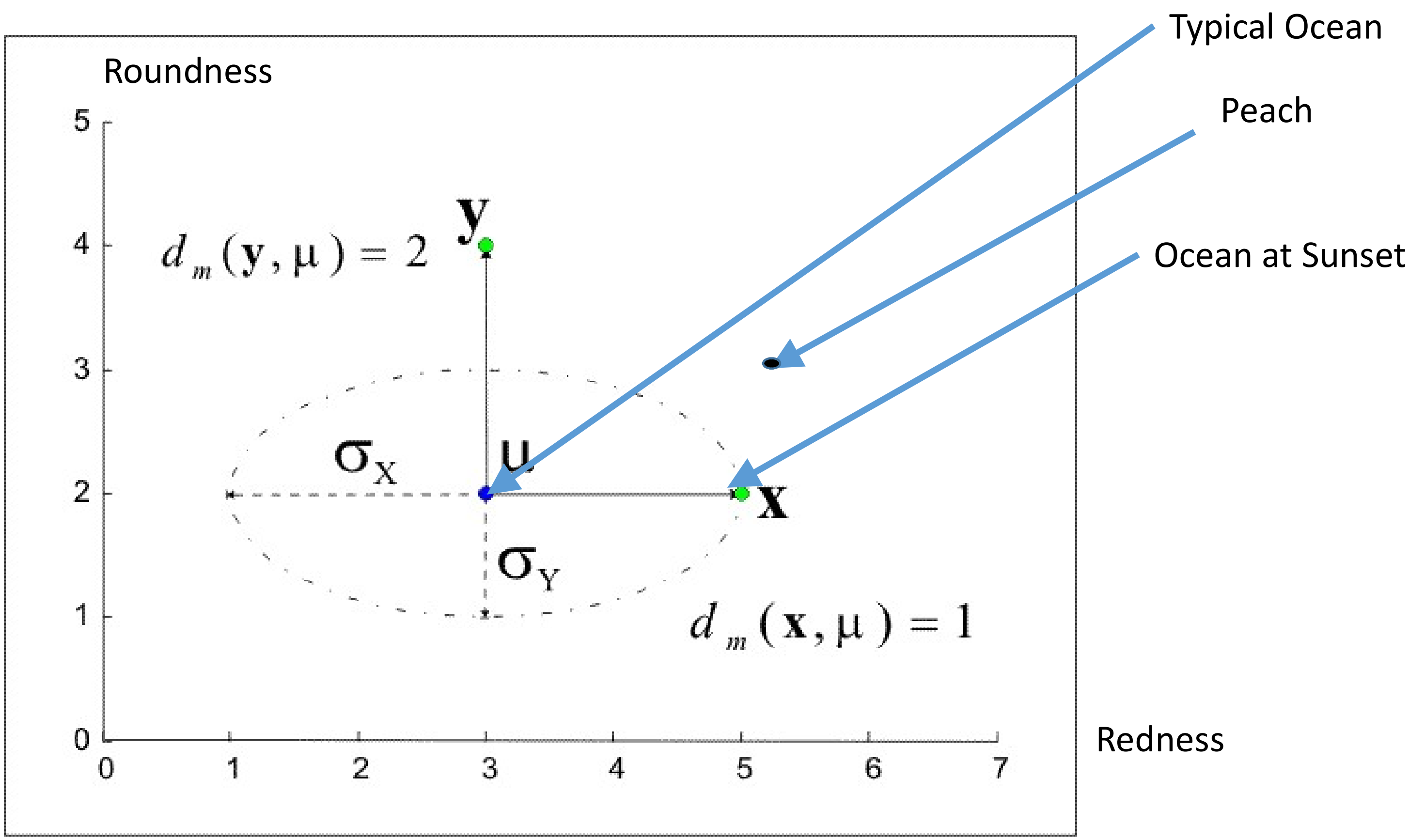
Today's Lecture

- Similarity and Dissimilarity of vectors: all you need is a norm
- Example: the Minkowski Norm (Lp norm)
- Cosine Similarity: you need a dot product
- Example: Diagonal Mahalanobis Distance
- What is Similarity?
- Metric Learning

What is similarity?



What is similarity?



Today’s Lecture

- Similarity and Dissimilarity of vectors: all you need is a norm
- Example: the Minkowski Norm (Lp norm)
- Cosine Similarity: you need a dot product
- Example: Diagonal Mahalanobis Distance
- What is Similarity?
- Metric Learning

Metric Learning

The goal: learn a function f(x,y) such that, if the user says y1 is more like x and y2 is less like x, then

$$f(x,y1) < f(x,y2)$$

a matrix W that parametrizes the similarity function $f_W(x, z) = x^T W z$.

Metric learning [edit]

Similarity learning is closely related to *distance metric learning*. Metric learning is the task of learning a distance function over objects. A metric or distance function has to obey four axioms: non-negativity, Identity of indiscernibles, symmetry and subadditivity / triangle inequality. In practice, metric learning algorithms ignore the condition of identity of indiscernibles and learn a pseudo-metric.

When the objects x_i are vectors in R^d, then any matrix W in the symmetric positive semi-definite cone S_+^d defines a distance pseudo-metric of the space of x through the form $D_W(x_1, x_2)^2 = (x_1 - x_2)^\top W(x_1 - x_2)$. When W is a symmetric positive definite matrix, D_W is a metric. Moreover, as any symmetric positive semi-definite matrix $W \in S_+^d$ can be decomposed as $W = L^\top L$ where $L \in R^{e \times d}$ and $e \geq rank(W)$, the distance function D_W can be rewritten equivalently $D_W(x_1, x_2)^2 = (x_1 - x_2)^\top L^\top L(x_1 - x_2) = \|L(x_1 - x_2)\|_2^2$. The distance $D_W(x_1, x_2)^2 = \|x_1' - x_2'\|_2^2$ corresponds to the Euclidean distance between the projected feature vectors $x_1' = Lx_1$ and $x_2' = Lx_2$. Some well-known approaches for metric learning include Large margin nearest neighbor,[4] Information theoretic metric learning (ITML).[5]

In statistics, the covariance matrix of the data is sometimes used to define a distance metric called Mahalanobis distance.

Applications [edit]

Similarity learning is used in information retrieval for learning to rank, in face verification or face identification,[6][7] and in recommendation systems. Also, many machine learning approaches rely on some metric. This includes unsupervised learning such as clustering, which groups together close or similar objects. It also includes supervised approaches like K-nearest neighbor algorithm which rely on labels of nearby objects to decide on the label of a new object. Metric learning has been proposed as a preprocessing step for many of these approaches .[8]

Scalability [edit]

Mahalanobis Distance Learning

- The goal is just to learn the parameters Σ so that

$$(\vec{x}-\vec{y})^T\Sigma^{-1}(\vec{x}-\vec{y}) = \frac{(x_1-y_1)^2}{\sigma_1^2} + \ldots + \frac{(x_D-y_D)^2}{\sigma_D^2}$$

accurately describes the perceived distance between x and y.

Sample problem

- Suppose your expriments show that people completely ignore dimension i. What should be the learned parameter σ_i^2?
- Suppose that dimension j is more important than dimension k. Should you have $\sigma_j^2 < \sigma_k^2$, or $\sigma_j^2 > \sigma_k^2$?
- Suppose that, instead of the normal Mahalanobis distance definition, you read a paper that does distance learning with

$$(\vec{x} - \vec{y})^T W(\vec{x} - \vec{y}) = w_1(x_1 - y_1)^2 + \cdots + w_D(x_D - y_D)^2$$

What's the relationship between the parameters w_j and σ_j^2?