Nearest Neighbor Classifiers

Hal Daumé III

CS5350: Machine Learning

04 Sept 2008

Today's Topics

1. The geometric view of data
2. Dealing with categorical values
3. One-nearest-neighbor classification
4. K-nearest-neighbor classification
5. Extensions...
6. Decision boundaries

The geometric view of data

We've already talked about representing a data point as a vector:

- The length of the vector is the total number of features
- Each element in the vector is the associated feature value

Once we've done this, we can think about data as points in high-dimensional space.

This enables us to use tools from linear algebra to think about learning problems:

- Distance between examples $\Longleftarrow$ this lecture
- Linear transformations
- Projections onto subspaces

Vectors and Matrices

We will often treat a data point as a vector. If there are D-many features, then $\boldsymbol{x} = \langle x_1, \dots, x_D \rangle$ can represent one example.

A data set of N examples can be represented as a matrix $\boldsymbol{X}$ defined so that $X_{n,d}$ is the dth feature value for the nth example.

Some things that are useful to know about vectors and matrices:

- Addition and subtraction of vectors/matrices is element-wise
- The norm of a vector is its length: $||\boldsymbol{x}|| = \sqrt{\sum_{d=1}^{D} x_d}$
- The dot product of two vectors is the sum of the product of their components: $(\boldsymbol{x} \cdot \boldsymbol{y}) = \sum_{d=1}^{D} x_d y_d$
  A dot product can be thought of as a measure of similarity. Two orthogonal vectors have a dot product of zero. Two parallel ones have a really high dot product.
- The norm can be expressed as $||\boldsymbol{x}|| = \sqrt{\boldsymbol{x} \cdot \boldsymbol{x}}$

Vectors and Matrices, continued...

Matrices with M rows and N columns are written as $\boldsymbol{A} \in \mathbb{R}^{M \times N}$.

- $\boldsymbol{A} \in \mathbb{R}^{L \times M}$ can be multiplied by $\boldsymbol{B} \in \mathbb{R}^{M \times N}$ to yield $\boldsymbol{C} \in \mathbb{R}^{L \times N}$. Elements are defined by: $C_{l,n} = \sum_m A_{l,m} B_{m,n}$
- The identity matrix is square with zeros everywhere except alone the diagonal, which has ones. It is denoted $\boldsymbol{I}$, or sometimes $\boldsymbol{I}_D$ to denote the dimensionality. In matlab, this is `eye(D)`.
- The transpose operator (eg., $A^\top$) flips a $M \times N$ matrix into a $N \times M$ matrix. In matlab, this is `A'`.
- So if a vector $\boldsymbol{x}$ is a $D \times 1$ matrix, then $\boldsymbol{x}^\top \boldsymbol{x}$ is a scalar, but $\boldsymbol{x}\boldsymbol{x}^\top$ is a matrix.

Reminder: Euclidean distance

Given $\boldsymbol{x} = \langle x_1, x_2, \dots, x_D \rangle$ and $\boldsymbol{y} = \langle y_1, y_2, \dots, y_D \rangle$, the Euclidean distance between $\boldsymbol{x}$ and $\boldsymbol{y}$ is defined by:

$$\begin{aligned} ||\boldsymbol{x} - \boldsymbol{y}|| &= \sqrt{\sum_{d=1}^{D} (x_d - y_d)^2} \\ &= \sqrt{\sum_{d=1}^{D} \left(x_d^2 + y_d^2 - 2x_d y_d\right)} \\ &= \sqrt{\sum_{d=1}^{D} x_d^2 + \sum_{d=1}^{D} y_d^2 - 2\sum_{d=1}^{D} x_d y_d} \\ &= \sqrt{||\boldsymbol{x}||^2 + ||\boldsymbol{y}||^2 - 2\boldsymbol{x} \cdot \boldsymbol{y}} \end{aligned}$$

Embedding the play/no-play data

Y	Out	T	R
P	Sunny	Low	Yes
N	Sunny	High	Yes
N	Sunny	High	No
P	Overcast	Low	Yes
P	Overcast	High	No
P	Overcast	Low	No
N	Rainy	Low	Yes
P	Rainy	Low	No

$\Rightarrow$

Y	⟨Out , T , R⟩
1	⟨ ? , 0 , 1 ⟩
0	⟨ ? , 1 , 1 ⟩
0	⟨ ? , 1 , 0 ⟩
1	⟨ ? , 0 , 1 ⟩
1	⟨ ? , 1 , 0 ⟩
1	⟨ ? , 0 , 0 ⟩
0	⟨ ? , 0 , 1 ⟩
1	⟨ ? , 0 , 0 ⟩

Why not just map “Sunny” to 0, “Overcast” to 1 and “Rainy” to 2?

Dealing with categorical values

Solution: map a categorical feature with K values into K binary features.

Y	Out	T	R
P	Sunny	Low	Yes
N	Sunny	High	Yes
N	Sunny	High	No
P	Overcast	Low	Yes
P	Overcast	High	No
P	Overcast	Low	No
N	Rainy	Low	Yes
P	Rainy	Low	No

$\Rightarrow$

Y	⟨S? , O? , R? , T , R⟩
1	⟨ 1 , 0 , 0 , 0 , 1 ⟩
0	⟨ 1 , 0 , 0 , 1 , 1 ⟩
0	⟨ 1 , 0 , 0 , 1 , 0 ⟩
1	⟨ 0 , 1 , 0 , 0 , 1 ⟩
1	⟨ 0 , 1 , 0 , 1 , 0 ⟩
1	⟨ 0 , 1 , 0 , 0 , 0 ⟩
0	⟨ 0 , 0 , 1 , 0 , 1 ⟩
1	⟨ 0 , 0 , 1 , 0 , 0 ⟩