Constructing the toy dataset

(see Section 3.1 in the paper)

The tasks were chosen as binary decision rules which are invariant under $O(3)$ rotations of the sphere, with 12 binary inputs that represent 12 uniformly distributed points on a 2D sphere; with such rules, 4096 different patterns of the input X are divided into 64 disjoint orbits of the rotation group; these orbits form a minimal sufficient statistics (MSS) for spherically symmetric rules.

In other words:

- the learning problem is posed as a binary classification (i.e., there are two classes)
- the data set is artificial, and constructed as follows:

  - 12 points are uniformly distributed on a sphere (note: a “2D sphere” is a sphere in 3D space!)
  - each point is a two-state system, or bit, i.e. takes on one of the two possible values $\{0, 1\}$
  - a system of 12 bits has $2^{12} = 4096$ different configurations (i.e. microstates); however, not all of them are distinct according to the decision rules
  - the decision rules are defined to be invariant under $O(3)$ rotations of the sphere; therefore, configurations that can be mapped into each other through $O(3)$ rotations of the sphere are indistinguishable
  - thus, 4096 states are split into 64 disjoint groups (orbits of the rotation group); this partition represents a MSS of the input space X

To “see” these orbits, imagine a sphere inscribed in a regular dodecahedron (fig. 1), the faces of the dodecahedron indicating the position of the 12 points uniformly distributed on the sphere.
Consider a configuration of one bit having the value 1 and the other eleven bits having the value 0. There are 12 options to place the 1; however, all of these 12 configurations can be transformed into each other through $O(3)$ rotations – thus, they form one orbit.
The configurations having two bits in state 1 and ten bits in state 0 form more than one orbit: The two 1 can be placed either on two adjacent faces (and all such configurations form one orbit), or on two faces separated by one face (another orbit), or on two opposite faces with maximum distance (another orbit). Continuing the counting in this fashion, one will find 64 distinct orbits.

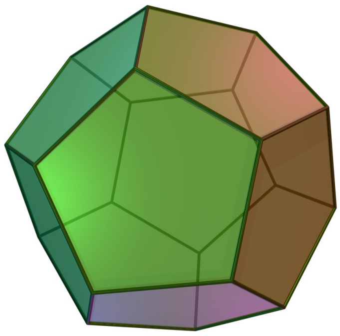

Figure 1: Dodecahedron. Image source: https://goo.gl/images/k6qOpw

- to generate the input-output distribution $P(X, Y)$, a spherically symmetric real-valued function of the pattern $f(x)$ was calculated (evaluated through its spherical harmonics power spectrum [Kazhdan et al. (2003)])

Brief summary of the referenced paper by Kazhdan et al.

Title: Rotation Invariant Spherical Harmonic Representations of 3D Shape Descriptors

Content: A new method for obtaining rotation-invariant representations of shapes in 3D is proposed.

Motivation: Data processing in many disciplines involves tools for acquiring and visualizing 3D models; as a result, retrieving models from large databases became a widespread need, calling for robust matching algorithms. The challenge in 3D shape matching arises from the fact that in many applications models should be considered the same if they differ by a rotation.

Method: The key idea of the new approach is to describe a spherical function in terms of the amount of energy it contains at different frequencies. Since these values do not change when the function is rotated, the resulting descriptor is rotation-invariant. This approach can be viewed as a generalization of the Fourier Descriptor method to the case of spherical functions.

Background:

- Spherical harmonics describe the way that rotations act on a spherical function. They can be used to represent a function on a sphere in a rotation-invariant manner.
- Any spherical function $f(\theta, \phi)$ can be decomposed as the sum of its harmonics:

$$f(\theta, \phi) = \sum_{l=0}^{\infty} \sum_{m=-l}^{m=l} a_{lm} Y_l^m(\theta, \phi) \tag{1}$$

- The key property of this decomposition: If we restrict to some frequency l, and define the subspace of functions

$$V_l = Span(Y_l^{-l}, Y_l^{-l+1}, ..., Y_l^{l-1}, Y_l^l), \tag{2}$$

then:

  * V_l is a representation for the rotation group: For any function $f \in V_l$ and any rotation R, we have $R(f) \in V_l$.

    In other words, if π_l is the projector onto the subspace V_l, then it commutes with rotations:

$$\pi_l(R(f)) = (R(\pi_l(f))) \tag{3}$$

  * V_l is irreducible: V_l can **not** be further decomposed as the direct sum $V_l = V_l' \oplus V_l''$, where V_l' and V_l'' are also non-trivial representations of the rotation group.

- Thus, the first property presents a way of decomposing spherical functions into rotation invariant components, while the second property guarantees that, in a linear sense, this decomposition is optimal.

Once this is figured out, we continue with obtaining the label and defining the conditional probability distribution:

- to obtain a binary label (i.e., $\{0, 1\}$) $y(x)$ for each pattern x, a step function was applied:

$$y(x) = \Theta(f(x) - \theta) \tag{4}$$

- the labels were then softened to a stochastic rule through a standard sigmoidal function $\psi(u) = 1/(1 + e^{-\gamma u})$; in other words, the conditional probability for a label y given x was defined as:

$$p(y = 1|x) = \psi(f(x) - \theta) \tag{5}$$

- the treshold θ was selected such that $p(y = 1) = \sum_x p(y = 1|x)p(x) \approx 0.5$, with uniform $p(x)$
- the sigmoidal gain γ was set to keep the mutual information at $I(X; Y) \approx 0.99$ bits