# What is K-Means?

**Color Extraction with K-Means in CIELAB Space**

Segmenting clothing items allows us to extract the color of each clothing item individually. In this context, a single dominant color is used to represent each item.

K-means is a cutting-edge approach for extracting colors from an image. It is a clustering method that groups a set of data points into "K" disjoint subsets, where points in each subset are close in distance. The K-means algorithm operates as follows:

1. Randomly select "k" points as the centroid of each subset. The choice of "k" is determined by the desired number of groups to separate the data.
2. Assign each data point to the closest group based on a certain metric (commonly Euclidean distance).
3. Recalculate the mean of points in each subset and set new centroids for each group. Reassign the data to the new clusters according to the new centroids.
4. Repeat the previous step until the model converges.

When applying the K-means algorithm to images, the first consideration is choosing an appropriate color space. Different color spaces, such as RGB, HSV, CMYK, and CIELAB, exist. In this project, we use the CIELAB (Lab) color space for clustering pixels due to its similarity to human vision, especially regarding the perception of lightness.

The Lab color model is a three-axis system. The L\* axis represents lightness vertically, ranging from darkest black at L\* = 0 to brightest white at L\* = 100. The a\* axis goes from green to red, while the b\* axis goes from blue to yellow. In practice, the values of a\* and b\* typically range from -128 to +127.

In our case, for each segmentation mask of clothing, we first map the mask to the original image to obtain RGB values of related pixels. We then transform the colors to Lab space and apply K-means to cluster pixels. Setting K = 5 to obtain five subsets, we choose the centroid of the subset with the maximum number of pixels as the dominant color of the segmented clothing. Finally, the centroid is transformed back to the RGB space for later processing.

**Figure: CIELAB color space**

A diagram of a color spectrum

Description automatically generated