Predicting Polar Bear Management Zones from Isotopic Features

Kevin Koehler

August 14, 2018

Abstract

We use various machine learning classifiers on isotopic composition features $\delta 2H, \delta 13C, \delta 15N$, and $\delta 18O$ to identify polar bear management zones.

1 Models

1.1 Bayesian Classifiers

1.1.1 Theory

Bayesian classifiers estimate the probability of a class label given a vector of features. Let $\mathbb{L} = \{l_1, ..., l_k\}$ be the set of class labels and each feature vector $\mathbf{x}$ be in $\mathbb{R}^n$, then the *maximum a posteriori (MAP)* classifier assigns the following label $\hat{y}$ to $\mathbf{x}$:

$$\hat{y} = \max_{l_i \in \mathbb{L}} P(l_i | x_1, ..., x_n)$$

A naïve Bayesan classifier reduces $P(l_i|x_1, ..., x_n)$ to $P(l_i) \prod_{x_j \in \mathbf{x}} P(x_j|l_i)$ by making the *naïve assumption* whereby we assume conditional independence between all $x_p, x_q \in \mathbf{x}$ where $p \neq q$. To see this, first the apply the chain rule to joint probability (which is proportional to the overall expression):

$$\begin{aligned} &P(c_i, x_1, ..., x_n) \\ =&P(x_1|x_2, ..., x_n, l_i)...P(x_{n-1}, x_n|l_i)P(x_n|l_i)P(l_i) \end{aligned}$$

With the naïve assumption we have $P(x_j|x_{j+1}, ..., x_n, l_i = P(x_j|l_i)$ thus every $P(x_j|x_{j+1}, ..., x_n, l_i) = P(x_j|l_i)$ yielding:

$$\begin{aligned}\hat{y} &= \prod_{x_j \in \mathbf{x}} P(x_j|l_i)P(l_i) \\ &= P(l_i) \prod_{x_j \in \mathbf{x}} P(x_j|l_i)\end{aligned}$$

1.1.2 Results

As is typical for problems of this nature, the naïve Bayesian classifier is surprisingly accurate. As one-hot encoding the features is not useful for Bayesian classifiers, the only pre-processing that was done was dropping incomplete feature vectors, as well as PCA, which had a negligible effect.

Since there is some randomness, the accuracy supplied is an average of 10 simulations. The accuracy for this model had a mean of 0.76281 with a variance of 0.0965.

For more advanced metrics, see the **Discussion** section of this paper.