The von Mises Naive Bayes Classifier for Angular Data. Appendix

Pedro L. López-Cruz, Concha Bielza, and Pedro Larrañaga

Computational Intelligence Group
Departamento de Inteligencia Artificial
Facultad de Informática
Universidad Politécnica de Madrid
Campus de Montegancedo
Boadilla del Monte, 28660, Madrid, SPAIN
Tel.: (+34)913363675
Fax: (+34)913524819
pedro.lcruz@upm.es,{mcbielza,pedro.larranaga}@fi.upm.es

1 Appendix

1.1 Decision Surfaces of the von Mises Naive Bayes Classifiers

One Predictive Variable. We assume that the class is binary, e.g., $val(C) = \{1, 2\}$. The classifier has one predictive variable Φ, with conditional probability distributions $(\Phi|C = c) \equiv \Phi^{(c)} \sim vM(\mu_\Phi^{(c)}, \kappa_\Phi^{(c)})$. We compute the decision surfaces induced by the von Mises NB classifiers by equaling the posterior probabilities of the class values as follows:

$$p(C = 1|\Phi = \phi) = p(C = 2|\Phi = \phi) \ . \tag{1}$$

Using Bayes' rule in (1), we get

$$p(C = 1)f_{\Phi|C=1}(\phi; \mu_\Phi^{(1)}, \kappa_\Phi^{(1)}) = p(C = 2)f_{\Phi|C=2}(\phi; \mu_\Phi^{(2)}, \kappa_\Phi^{(2)}) \ . \tag{2}$$

We substitute the von Mises probability density function in (2):

$$p(C = 1)\frac{1}{2\pi I_0(\kappa_\Phi^{(1)})} \exp(\kappa_\Phi^{(1)} \cos(\phi - \mu_\Phi^{(1)})) = p(C = 2)\frac{1}{2\pi I_0(\kappa_\Phi^{(2)})} \exp(\kappa_\Phi^{(2)} \cos(\phi - \mu_\Phi^{(2)})) \ .$$

Then we simplify the constant 2π, take logarithms and arrange all terms on the same side of the equation:

$$\kappa_\Phi^{(1)} \cos(\phi - \mu_\Phi^{(1)}) - \kappa_\Phi^{(2)} \cos(\phi - \mu_\Phi^{(2)}) + \ln \frac{p(C = 1)}{I_0(\kappa_\Phi^{(1)})} - \ln \frac{p(C = 2)}{I_0(\kappa_\Phi^{(2)})} = 0 \ .$$

We substitute $\cos(\beta - \gamma) = \cos(\beta)\cos(\gamma) + \sin(\beta)\sin(\gamma)$ and operate the logarithms

$$\kappa_{\Phi}^{(1)} \left[\cos\phi\cos\mu_{\Phi}^{(1)} + \sin\phi\sin\mu_{\Phi}^{(1)}\right] - \kappa_{\Phi}^{(2)} \left[\cos\phi\cos\mu_{\Phi}^{(2)} + \sin\phi\sin\mu_{\Phi}^{(2)}\right] + \ln\frac{p(C=1)I_0(\kappa_{\Phi}^{(2)})}{p(C=2)I_0(\kappa_{\Phi}^{(1)})} = 0,$$

which we arrange using $\cos\phi$ and $\sin\phi$ as common terms:

$$(\kappa_{\Phi}^{(1)}\cos\mu_{\Phi}^{(1)} - \kappa_{\Phi}^{(2)}\cos\mu_{\Phi}^{(2)})\cos\phi + (\kappa_{\Phi}^{(1)}\sin\mu_{\Phi}^{(1)} - \kappa_{\Phi}^{(2)}\sin\mu_{\Phi}^{(2)})\sin\phi + \ln\frac{p(C=1)I_0(\kappa_{\Phi}^{(2)})}{p(C=2)I_0(\kappa_{\Phi}^{(1)})} = 0 \ . \quad (3)$$

We substitute $a = \kappa_{\Phi}^{(1)}\cos\mu_{\Phi}^{(1)} - \kappa_{\Phi}^{(2)}\cos\mu_{\Phi}^{(2)}, b = \kappa_{\Phi}^{(1)}\sin\mu_{\Phi}^{(1)} - \kappa_{\Phi}^{(2)}\sin\mu_{\Phi}^{(2)}$, $D = -\ln\frac{p(C=1)I_0(\kappa_{\Phi}^{(2)})}{p(C=2)I_0(\kappa_{\Phi}^{(1)})}$, and get

$$a\cos\phi + b\sin\phi = D \ .$$

Trigonometrically, this is equivalent to

$$T\cos(\phi - \alpha) = D, \quad (4)$$

where $T = \sqrt{a^2 + b^2}, \cos\alpha = a/T, \sin\alpha = b/T, \tan\alpha = b/a$.

Isolating ϕ from (4), we get

$$\phi = \alpha \pm \arccos(D/T) \ . \quad (5)$$

The NB classifier finds two angles that bound the class regions.

To compute the decision surface in the Euclidean 2-dimensional space, we have to compute the corresponding Cartesian coordinates using the following expression:

$$(x, y) = (\cos\phi, \sin\phi) \ . \quad (6)$$

Substituting (6) in (3), we get

$$(\kappa_{\Phi}^{(1)}\mu_X^{(1)} - \kappa_{\Phi}^{(2)}\mu_X^{(2)})x + (\kappa_{\Phi}^{(1)}\mu_Y^{(1)} - \kappa_{\Phi}^{(2)}\mu_Y^{(2)})y + \ln\frac{p(C=1)I_0(\kappa_{\Phi}^{(2)})}{p(C=2)I_0(\kappa_{\Phi}^{(1)})} = 0, \quad (7)$$

where $(\mu_X^{(1)}, \mu_Y^{(1)})$ and $(\mu_X^{(2)}, \mu_Y^{(2)})$ are the Cartesian coordinates of the points on the circle defined by the mean directions $\mu_{\Phi}^{(1)}$ and $\mu_{\Phi}^{(2)}$, respectively.

Equation (7) defines a line that goes through the points specified by the decision angles induced by von Mises NB and derived in (5). Therefore, von Mises NB with one predictive variable is a linear classifier and should be especially well suited for solving problems with linearly separable classes.

Two Predictive Variables. In this scenario, we have two circular predictive variables Φ and Ψ. The domain defined by these variables is a torus $(-\pi, \pi] \times (-\pi, \pi]$. As in the simpler case above, we compute the decision surfaces induced by the classifier by equaling the posterior probability of the two class values

$$p(C = 1|\Phi = \phi, \Psi = \psi) = p(C = 2|\Phi = \phi, \Psi = \psi) \ .$$

Using Bayes' rule and the conditional independence assumption, we get

$$\begin{aligned} &p(C = 1) f_{\Phi|C=1}(\phi; \mu_\Phi^{(1)}, \kappa_\Phi^{(1)}) f_{\Psi|C=1}(\psi; \mu_\Psi^{(1)}, \kappa_\Psi^{(1)}) \\ &\qquad = p(C = 2) f_{\Phi|C=2}(\phi; \mu_\Phi^{(2)}, \kappa_\Phi^{(2)}) f_{\Psi|C=2}(\psi; \mu_\Psi^{(2)}, \kappa_\Psi^{(2)}) \ . \end{aligned} \tag{8}$$

We substitute the von Mises density in (8):

$$\begin{aligned} &p(C = 1) \frac{\exp(\kappa_\Phi^{(1)} \cos(\phi - \mu_\Phi^{(1)}))}{2\pi I_0(\kappa_\Phi^{(1)})} \frac{\exp(\kappa_\Psi^{(1)} \cos(\psi - \mu_\Psi^{(1)}))}{2\pi I_0(\kappa_\Psi^{(1)})} \\ &\qquad = p(C = 2) \frac{\exp(\kappa_\Phi^{(2)} \cos(\phi - \mu_\Phi^{(2)}))}{2\pi I_0(\kappa_\Phi^{(2)})} \frac{\exp(\kappa_\Psi^{(2)} \cos(\psi - \mu_\Psi^{(2)}))}{2\pi I_0(\kappa_\Psi^{(2)})} \ . \end{aligned}$$

We simplify the constant 2π, take logarithms and arrange all the terms on the same side of the equation:

$$\begin{aligned} &\kappa_\Phi^{(1)} \cos(\phi - \mu_\Phi^{(1)}) + \kappa_\Psi^{(1)} \cos(\psi - \mu_\Psi^{(1)}) \\ &\qquad - \kappa_\Phi^{(2)} \cos(\phi - \mu_\Phi^{(2)}) - \kappa_\Psi^{(2)} \cos(\psi - \mu_\Psi^{(2)}) \\ &\qquad\qquad + \ln \frac{p(C = 1) I_0(\kappa_\Phi^{(2)}) I_0(\kappa_\Psi^{(2)})}{p(C = 2) I_0(\kappa_\Phi^{(1)}) I_0(\kappa_\Psi^{(1)})} = 0 \ . \end{aligned}$$

We substitute the trigonometric identity $\cos(\beta - \gamma) = \cos(\beta)\cos(\gamma) + \sin(\beta)\sin(\gamma)$ and arrange the terms:

$$\begin{aligned} &(\kappa_\Phi^{(1)} \cos \mu_\Phi^{(1)} - \kappa_\Phi^{(2)} \cos \mu_\Phi^{(2)}) \cos \phi + (\kappa_\Phi^{(1)} \sin \mu_\Phi^{(1)} - \kappa_\Phi^{(2)} \sin \mu_\Phi^{(2)}) \sin \phi \\ &\quad + (\kappa_\Psi^{(1)} \cos \mu_\Psi^{(1)} - \kappa_\Psi^{(2)} \cos \mu_\Psi^{(2)}) \cos \psi + (\kappa_\Psi^{(1)} \sin \mu_\Psi^{(1)} - \kappa_\Psi^{(2)} \sin \mu_\Psi^{(2)}) \sin \psi \\ &\qquad + \ln \frac{p(C = 1) I_0(\kappa_\Phi^{(2)}) I_0(\kappa_\Psi^{(2)})}{p(C = 2) I_0(\kappa_\Phi^{(1)}) I_0(\kappa_\Psi^{(1)})} = 0 \ . \end{aligned} \tag{9}$$

We define the following constants:

$$\begin{aligned} a &= \kappa_\Phi^{(1)} \cos \mu_\Phi^{(1)} - \kappa_\Phi^{(2)} \cos \mu_\Phi^{(2)}, \\ b &= \kappa_\Phi^{(1)} \sin \mu_\Phi^{(1)} - \kappa_\Phi^{(2)} \sin \mu_\Phi^{(2)}, \\ c &= \kappa_\Psi^{(1)} \cos \mu_\Psi^{(1)} - \kappa_\Psi^{(2)} \cos \mu_\Psi^{(2)}, \\ d &= \kappa_\Psi^{(1)} \sin \mu_\Psi^{(1)} - \kappa_\Psi^{(2)} \sin \mu_\Psi^{(2)}, \\ D &= -\ln \frac{p(C = 1) I_0(\kappa_\Phi^{(2)}) I_0(\kappa_\Psi^{(2)})}{p(C = 2) I_0(\kappa_\Phi^{(1)}) I_0(\kappa_\Psi^{(1)})}, \end{aligned}$$

and substitute them in (9) to get

$$a \cos \phi + b \sin \phi + c \cos \psi + d \sin \psi = D \ . \tag{10}$$

The Cartesian coordinates of the points defined by the angles ϕ and ψ on the surface of a torus are

$$\begin{aligned} x &= (L + l \cos \phi) \cos \psi, \\ y &= (L + l \cos \phi) \sin \psi, \\ z &= l \sin \phi, \end{aligned}$$

where L is the distance from the center of the torus to the center of the revolving circumference that generates the torus, and l is the radius of the revolving circumference. We isolate the trigonometric functions and get

$$\begin{aligned} \sin \phi &= z/l, \\ \cos \phi &= \pm\sqrt{1 - \sin^2 \phi} = \pm\sqrt{1 - \left(\frac{z}{l}\right)^2} = \pm\frac{1}{l}\sqrt{l^2 - z^2}, \\ \sin \psi &= \frac{y}{L + l \cos \phi}, \\ \cos \psi &= \frac{x}{L + l \cos \phi} \ . \end{aligned}$$

Substituting these expressions in (10), we get the two following equations corresponding to the two signs of $\cos \phi$:

$$\begin{aligned} \frac{a}{l}\sqrt{l^2 - z^2} + \frac{b}{l}z + \frac{c}{L + \sqrt{l^2 - z^2}}x + \frac{d}{L + \sqrt{l^2 - z^2}}y + D = 0, \\ -\frac{a}{l}\sqrt{l^2 - z^2} + \frac{b}{l}z + \frac{c}{L - \sqrt{l^2 - z^2}}x + \frac{d}{L - \sqrt{l^2 - z^2}}y + D = 0 \ . \end{aligned}$$

Operating and arranging the terms, we get

$$\begin{aligned} clx + dly - az^2 + bz\sqrt{l^2 - z^2} + bLz + (aL + Dl)\sqrt{l^2 - z^2} + al^2 + DLl = 0, \\ clx + dly - az^2 - bz\sqrt{l^2 - z^2} + bLz - (aL + Dl)\sqrt{l^2 - z^2} + al^2 + DLl = 0 \ . \end{aligned} \tag{11}$$

The expressions in (11) are quadratic in z. Therefore, we conclude that von Mises NB with two predictive variables is a much more complex and flexible classifier than von Mises NB with one predictive variable.