Gene Earth Documentation

Min Hyung (Daniel) Kang

July 2015

1. To evaluate the success of a recently released movie a survey has been conducted: 1000 people were asked if they like the movie. The data are in file movieFINAL.txt (y = 1: like the movie; y = 0: do not like the movie). Besides, the information about age and viewers education was obtained (ed=0:no hight school; ed=1: high school; ed=2: BC degree; ed=3: MS degree or higher). Use read.table with option header=T to download the data.

2. Use education status as a continuous variable (as is) and as a dummy variable to set up and run the logistic regression model. Test the hypothesis that the dummy variable approach is equivalent to treating education status continuosly by testing $B_1 - B_0 = B_2 - B_1 = B_3 - B_2$ by likelihood-ratio test (B_0 is the coefficient at no high school, B_1 is the coefficient at high school, etc.).

3. Plot y versus age for viewers with high school degree and superimpose with the fitted model values.

4. Estimate the proportion of people who liked the movie on average. Compute this proportion using two methods: (1) using y observations; (2) using the estimated model. Do the results match?

5. What is the chance that a 32 years old person with a college degree likes the movie? Compute the 95% CI for this proportion on the logit scale and then transform to the probability scale.

6. Compute the p-value for the null hypothesis that people with BC and MS education of the same age equally like the movie. Use the Wald test.

1 Building GeneEarth

1.1 Projection

Assume we have n points on the sphere of unit radius, which are defined by angles α and β.

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} \cos\alpha\cos\beta \\ \sin\alpha\cos\beta \\ \sin\beta \end{bmatrix} \tag{1}$$

Now let camera be positioned with its direction of view defined by $\alpha_{view}, \beta_{view}$,

$$\begin{bmatrix} x_{view} \\ y_{view} \\ z_{view} \end{bmatrix} = \begin{bmatrix} \cos\alpha_{view}\cos\beta_{view} \\ \sin\alpha_{view}\cos\beta_{view} \\ \sin\beta_{view} \end{bmatrix} \tag{2}$$

Then projection of vector(1) on the plane orthogonal to (2) is defined as follows :

$$\begin{bmatrix} \cos\alpha_{view}\sin\beta_{view} & \sin\alpha_{view}\sin\beta_{view} & -\cos\beta_{view} \\ \sin\alpha_{view} & -\cos\alpha_{view} & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix}$$

$$= \begin{bmatrix} x\cos\alpha_{view}\sin\beta_{view} + y\sin\alpha_{view}\sin\beta_{view} - z\cos\beta_{view} \\ x\sin\alpha_{view} - y\cos\alpha_{view} \end{bmatrix} \tag{3}$$

2 Methods

2.1 Locating Gene

The user has the choice of locating a gene on the gene earth by using the right search panel, implemented by the method *centerLocation(x,y,z)* called by *goToGene*. Refer to equation (3). We note that to center the genes, $x \cos \alpha_{view} \sin \beta_{view} + y \sin \alpha_{view} \sin \beta_{view} - z \cos \beta_{view} = 0$ and $x \sin \alpha_{view} - y \cos \alpha_{view} = 0$.

1. $x = 0$
   Since $x \sin \alpha_{view} - y \cos \alpha_{view} = -y \cos \alpha_{view} = 0$, let $\alpha_{view} = \frac{\pi}{2}$
   Then $x \cos \alpha_{view} \sin \beta_{view} + y \sin \alpha_{view} \sin \beta_{view} - z \cos \beta_{view} = y \sin \beta_{view} - z \cos \beta_{view} = 0$
   Hence, $y \sin \beta_{view} = z \cos \beta_{view}$

   (a) $(y = 0)$ $\beta_{view} = \frac{\pi}{2}$

   (b) (else) $\frac{z}{y} = \frac{\cos \beta_{view}}{\sin \beta_{view}} = \tan \beta_{view}$. And hence, $\beta_{view} = \tan^{-1} \frac{z}{y}$

2. $x \neq 0$
   Since $x \sin \alpha_{view} - y \cos \alpha_{view} = 0$, $x \sin \alpha_{view} = y \cos \alpha_{view}$ we know $\alpha_{view} = \tan^{-1} \frac{y}{x}$

   Since $x \cos \alpha_{view} \sin \beta_{view} + y \sin \alpha_{view} \sin \beta_{view} - z \cos \beta_{view} = (x \cos \alpha_{view} + y \sin \alpha_{view}) \sin \beta_{view} = z \cos \beta_{view}$

   (a) $(x \cos \alpha_{view} + y \sin \alpha_{view} = 0)$ $\cos \beta_{view} = 0$ and $\beta_{view} = \frac{\pi}{2}$

   (b) (else) $\frac{z}{x \cos \alpha_{view} + y \sin \alpha_{view}} = \frac{\sin \beta_{view}}{\cos \beta_{view}} = \tan \beta_{view}$ and so $\beta_{view} = \tan^{-1} \frac{z}{x \cos \alpha_{view} + y \sin \alpha_{view}}$

   Note, however, that $\tan^{-1}$ only ranges from $-\frac{\pi}{2}$ to $\frac{\pi}{2}$. To take that into account, using the values of $\alpha_{view}, \beta_{view}$ we get from above computation, we calculate the condition value $(x_{view} * x + y_{view} * y + z_{view} * z)$ and adjust β_{view} value accordingly to make the gene face forward, not backward in the globe.

2.2 Zooming on a gene

Upon double clicking on a arbitrary spot, the user can make the program center on the location. We achieve this in following steps

1. Calculate the distance from the center of the globe, and compute the x,y coordinate of that location. (Check if the new location is within the globe)

   Let's denote the clicked location to be $(e.getX(), e.getY())$. Note that this is the coordinate within the panel, but in the program we have translation as well as zoom. Hence, to get the relative location, we compute as following :

   $$(x_{new}, y_{new}) = \left((e.getX() - translateX) * \frac{dist}{MULTIPLIER}, (e.getY() - translateY) * \frac{dist}{MULTIPLIER} \right)$$

   We can check whether the new location is outside the circle by simple calculation :

   $$x_{new}^2 + y_{new}^2 \leq 1$$

2. Using current $alpha_{view}$ and $beta_{view}$, compute the 3d coordinates of the clicked location.

We use equation(3) as well as the fact that the point is on sphere of unit radius to inversely calculate 3d-coordinate x, y, z

$$\text{Let } \begin{cases} x_{new} & = x\cos\alpha_{view}\sin\beta_{view} + y\sin\alpha_{view}\sin\beta_{view} - z\cos\beta_{view} \\ y_{new} & = x\sin\alpha_{view} - y\cos\alpha_{view} \\ 1 & = x^2 + y^2 + z^2 \end{cases}$$

1. We can make these equations as :

$$\begin{cases} 1) x_{new}sin\alpha_{view} & = x\cos\alpha_{view}\sin\alpha_{view}\sin\beta_{view} + y\sin^2\alpha_{view}\sin\beta_{view} - z\sin\alpha_{view}\cos\beta_{view} \\ 2) y_{new}\cos\alpha_{view}\sin\beta_{view} & = x\cos\alpha_{view}\sin\alpha_{view}\sin\beta_{view} - y\cos^2\alpha_{view}\sin\beta_{view} \end{cases}$$

Subtracting 2nd equation from 1st gives us :

$$x_{new}sin\alpha_{view} - y_{new}\cos\alpha_{view}\sin\beta_{view} = y\sin\beta_{view}(\sin^2\alpha_{view} + \cos^2\alpha_{view}) - z\sin\alpha_{view}\cos\beta_{view}$$
$$= y\sin\beta_{view} - z\sin\alpha_{view}\cos\beta_{view}$$
$$y = \frac{x_{new}sin\alpha_{view} - y_{new}\cos\alpha_{view}\sin\beta_{view} + z\sin\alpha_{view}\cos\beta_{view}}{\sin\beta_{view}}$$

2. We can also make original equations as :

$$\begin{cases} 1) x_{new}\cos\alpha_{view} & = x\cos^2\alpha_{view}\sin\beta_{view} + y\cos\alpha_{view}\sin\alpha_{view}\sin\beta_{view} - z\cos\alpha_{view}\cos\beta_{view} \\ 2) y_{new}\sin\alpha_{view}\sin\beta_{view} & = x\sin^2\alpha_{view}\sin\beta_{view} - y\cos\alpha_{view}\sin\alpha_{view}\sin\beta_{view} \end{cases}$$

Adding 1st and 2nd equations gives us :

$$x_{new}\cos\alpha_{view} + y_{new}\sin\alpha_{view}\sin\beta_{view} = x\sin\beta_{view}(\cos^2\alpha_{view} + \sin^2\alpha_{view}) - z\cos\alpha_{view}\cos\beta_{view}$$
$$= x\sin\beta_{view} - z\cos\alpha_{view}\cos\beta_{view}$$
$$x = \frac{x_{new}\cos\alpha_{view} + y_{new}\sin\alpha_{view}\sin\beta_{view} + z\cos\alpha_{view}\cos\beta_{view}}{\sin\beta_{view}}$$

3. We know that $x^2 + y^2 + z^2 = 1$

   (a) Let : $\sin\beta_{view}^2(x^2 + y^2 + z^2) = \sin\beta_{view}^2$

   $\sin\beta_{view}^2 = z^2\sin\beta_{view}^2 + x_{new}^2\cos\alpha_{view}^2 + y_{new}^2\sin\alpha_{view}^2\sin\beta_{view}^2 + z^2\cos\alpha_{view}^2\cos\beta_{view}^2$
   $+ 2(x_{new}y_{new}\cos\alpha_{view}\sin\alpha_{view}\sin\beta_{view} + x_{new}z\cos\alpha_{view}^2\cos\beta_{view} + y_{new}z\cos\alpha_{view}\sin\alpha_{view}\cos\beta_{view}\sin\beta_{view})$
   $+ x_{new}^2\sin\alpha_{view}^2 + y_{new}^2\cos\alpha_{view}^2\sin\beta_{view}^2 + z^2\sin\alpha_{view}^2\cos\beta_{view}^2$
   $+ 2(-x_{new}y_{new}\cos\alpha_{view}\sin\alpha_{view}\sin\beta_{view} + x_{new}z\sin\alpha_{view}^2\cos\beta_{view} - y_{new}z\cos\alpha_{view}\sin\alpha_{view}\cos\beta_{view}\sin\beta_{view})$
   $= z^2\sin\beta_{view}^2 + x_{new}^2 + y_{new}^2\sin\beta_{view}^2 + z^2\cos\beta_{view}^2 + 2x_{new}z\cos\beta_{view}$
   $= x_{new}^2 + y_{new}^2\sin\beta_{view}^2 + z^2 + 2x_{new}z\cos\beta_{view}$

4

Hence we get the following quadratic equation with respect to z :

$$z^2 + 2x_{new}\cos\beta_{view}z + (x_{new}^2 + (y_{new}^2 - 1)\sin\beta_{view}^2) = 0$$

We compute the value of z using quadratic equation, and plug in to the values of equations we got above to get values of x, y, z.

(b) Note, that when $\sin\beta_{view} = 0$ above definition is undefined. Hence, we compute for such special case.

$$\text{We have } \begin{cases} x_{new} &= -z\cos\beta_{view} \\ y_{new} &= x\sin\alpha_{view} - y\cos\alpha_{view} \\ 1 &= x^2 + y^2 + z^2 \end{cases}$$

We then have 1. $z = \frac{-x_{new}}{\cos\beta_{view}}$ 2. $y = \frac{x\sin\alpha_{view} - y_{new}}{\cos\alpha_{view}}$. We plug this into third equation and multiply both sides by $\cos\alpha_{view}^2\cos\beta_{view}^2$.

$$\cos\alpha_{view}^2\cos\beta_{view}^2 = \cos\alpha_{view}^2\cos\beta_{view}^2(x^2 + (\frac{x\sin\alpha_{view} - y_{new}}{\cos\alpha_{view}})^2 + (\frac{-x_{new}}{\cos\beta_{view}})^2)$$
$$= \cos\alpha_{view}^2\cos\beta_{view}^2x^2 + \cos\beta_{view}^2(\sin\alpha_{view}^2x^2 - 2sin\alpha_{view}y_{new}x + y_{new}^2) + \cos\alpha_{view}^2x_{new}^2$$
$$0 = \cos\beta_{view}^2x^2 - 2sin\alpha_{view}\cos\beta_{view}^2y_{new}x + (\cos\beta_{view}^2y_{new}^2 + \cos\alpha_{view}^2x_{new}^2 - \cos\alpha_{view}^2\cos\beta_{view}^2)$$

We use quadratic equation to get value of x, corresponding y and z.

(c) Last special case is if $\cos\alpha_{view} = 0$ as well. Then we have :

$$\begin{cases} x_{new} &= -z\cos\beta_{view} \\ y_{new} &= x\sin\alpha_{view} \\ 1 &= x^2 + y^2 + z^2 \end{cases}$$

Hence, we have $x = \frac{y_{new}}{\sin\alpha_{view}}, z = \frac{-x_{new}}{\cos\beta_{view}}, y = \pm\sqrt{1 - z^2 - x^2}$

For all of the above computations, since we use quadratic equation, we can have different values. We test if we have right value by checking the $cond = x_{view} * x + y_{view} * y + z_{view} * z$, and if it is positive.

3. Using $centerLocation(x,y,z)$, we compute a new $alpha_{view}$ and $beta_{view}$ that will allow us to center around that location.

2.3 Rotation

We can use the method used for zooming on a gene to rotate. When user clicks on a point 1 and drags to point 2, we calculate as if point 2 is the center and point 1 was double clicked. That is,

$$\begin{cases} dx &= (x_1 - x_2) * \frac{dist}{multiplier} \\ dy &= (y_1 - y_2) * \frac{dist}{multiplier} \end{cases}$$

and convert (dx, dy) to 3-d coordinates(as in 2.2), and we center the location on that coordinate(as in 2.1).

We check for two cases in rotation : 1. Whether dx, dy is significant enough, for mouse movement is subtle and could lead to undesired calculations and movements.

5