

Computational Neuroscience - Homework Module 2

Sepehr SAEEDPOUR

May 2025

1 Covariance and Correlation

1.1 Variance of firing rates

We need to show that $\text{Var}(r_1) = \frac{1}{N-1} \|x\|^2$.

The variance of r_1 is defined as:

$$\text{Var}(r_1) = \frac{1}{N-1} \sum_{i=1}^N (r_{1,i} - \bar{r}_1)^2 \quad (1)$$

The squared norm of x is:

$$\|x\|^2 = (r_{1,1} - \bar{r}_1)^2 + (r_{1,2} - \bar{r}_1)^2 + \dots + (r_{1,N} - \bar{r}_1)^2 \quad (2)$$

$$= \sum_{i=1}^N (r_{1,i} - \bar{r}_1)^2 \quad (3)$$

Therefore:

$$\text{Var}(r_1) = \frac{1}{N-1} \sum_{i=1}^N (r_{1,i} - \bar{r}_1)^2 \quad (4)$$

$$= \frac{1}{N-1} \|x\|^2 \quad (5)$$

Please note that when we estimate variance from a sample using the sample mean, the result tends to underestimate the true population variance (i.e., biased estimator). $N-1$ as a divisor instead N (Bessel's correction) fixes this by compensating for the loss of one degree of freedom from using the sample mean. This adjustment makes the sample variance an unbiased estimator of the population variance.

1.2 Cosine of angle between x and y

The cosine of the angle between two vectors x and y is:

$$\cos(\theta) = \frac{x \cdot y}{\|x\| \cdot \|y\|} \quad (6)$$

Let's calculate each component:

$$x \cdot y = \sum_{i=1}^N (r_{1,i} - \bar{r}_1)(r_{2,i} - \bar{r}_2) \quad (7)$$

$$= (N-1) \cdot \text{Cov}(r_1, r_2) \quad (8)$$

For the denominator:

$$\|x\| \cdot \|y\| = \sqrt{\sum_{i=1}^N (r_{1,i} - \bar{r}_1)^2} \cdot \sqrt{\sum_{i=1}^N (r_{2,i} - \bar{r}_2)^2} \quad (9)$$

$$= \sqrt{(N-1) \cdot \text{Var}(r_1)} \cdot \sqrt{(N-1) \cdot \text{Var}(r_2)} \quad (10)$$

$$= (N-1) \cdot \sqrt{\text{Var}(r_1) \cdot \text{Var}(r_2)} \quad (11)$$

Therefore:

$$\cos(\theta) = \frac{(N-1) \cdot \text{Cov}(r_1, r_2)}{(N-1) \cdot \sqrt{\text{Var}(r_1) \cdot \text{Var}(r_2)}} \quad (12)$$

$$= \frac{\text{Cov}(r_1, r_2)}{\sqrt{\text{Var}(r_1) \cdot \text{Var}(r_2)}} \quad (13)$$

$$= \text{corr}(r_1, r_2) \quad (14)$$

Thus, the cosine of the angle between x and y is exactly the correlation coefficient between r_1 and r_2 .

1.3 Maximum and minimum values of correlation coefficient

The correlation coefficient (and therefore the cosine of the angle) ranges from -1 to 1 :

- Maximum value: 1 (when the vectors are perfectly aligned, i.e., angle = 0°)
- Minimum value: -1 (when the vectors point in opposite directions, i.e., angle = 180°)

The correlation is 1 when the two variables are perfectly positively linearly related, and -1 when they are perfectly negatively linearly related. The correlation is 0 when there is no linear relationship between the variables.

1.4 Meaning of "centered"

Here, we're centering the firing rates by subtracting their respective means (\bar{r}_1 and \bar{r}_2). In fact, "Centered" refers to subtracting the mean from each data point, moving the mean of the data to zero. This transformation moves the data to be centered around the origin in the coordinate system. It's an important preprocessing step in many statistical analyses (e.g., PCA)

2 Ridge Regression – Closed-Form Solution

2.1 Penalizing complex models

The term $\lambda \|w\|^2$ penalizes complex models by adding a cost proportional to the squared magnitude of the weight vector. This regularization:

- Discourages large weights, which often correspond to overfitting
- Makes the model less sensitive to noise in the training data
- Promotes smoother, more generalizable models
- Stabilizes the solution when features are highly correlated

Intuitively, large weights amplify the influence of individual features, potentially allowing the model to fit noise rather than the underlying pattern. By keeping the weights small, the model becomes more robust.

2.2 Solution for $\lambda = 0$ (no regularization)

Starting with the error function:

$$E(w) = \|y - Xw\|^2 \quad (15)$$

Expanding:

$$E(w) = (y - Xw)^T (y - Xw) \quad (16)$$

$$= y^T y - y^T Xw - w^T X^T y + w^T X^T Xw \quad (17)$$

$$= y^T y - 2w^T X^T y + w^T X^T Xw \quad (\text{since } y^T Xw \text{ is a scalar}) \quad (18)$$

To find the minimum, we take the derivative with respect to w and set it to zero:

$$\frac{\partial E}{\partial w} = -2X^T y + 2X^T X w = 0 \quad (19)$$

Solving for w :

$$X^T X w = X^T y \quad (20)$$

$$w = (X^T X)^{-1} X^T y \quad (21)$$

This is the normal equation solution, where $(X^T X)^{-1} X^T$ is the pseudo-inverse of X .

2.3 Solution for $\lambda > 0$ (with regularization)

With regularization:

$$E(w) = \|y - Xw\|^2 + \lambda \|w\|^2 \quad (22)$$

$$= (y - Xw)^T (y - Xw) + \lambda w^T w \quad (23)$$

$$= y^T y - 2w^T X^T y + w^T X^T X w + \lambda w^T w \quad (24)$$

$$= y^T y - 2w^T X^T y + w^T (X^T X + \lambda I) w \quad (25)$$

Taking the derivative and setting it to zero:

$$\frac{\partial E}{\partial w} = -2X^T y + 2(X^T X + \lambda I)w = 0 \quad (26)$$

Solving for w :

$$(X^T X + \lambda I)w = X^T y \quad (27)$$

$$w = (X^T X + \lambda I)^{-1} X^T y \quad (28)$$

2.4 Purpose of adding λI to $X^T X$

Adding λI to $X^T X$ addresses potential numerical instability and invertibility issues. Specifically:

- It ensures that $X^T X + \lambda I$ is always invertible, even when $X^T X$ is singular or near-singular
- It provides a solution when features are highly correlated (multicollinearity)
- It improves the condition number of the matrix, making the solution more stable
- It enables a solution even when there are more features than data points ($d > n$)

Without regularization, $X^T X$ might be rank-deficient (i.e., some eigenvalues are zero), making inversion impossible. Since λI has all eigenvalues equal to $\lambda > 0$, adding it to $X^T X$ ensures all eigenvalues of the resulting matrix are positive, making it positive definite and therefore invertible.

2.5 Optimal value for λ

There is no single "optimal" value for λ that works for all problems. The optimal λ is typically determined through cross-validation:

1. Split the data into training and validation sets
2. Fit the model on the training set with different values of λ
3. Evaluate the model's performance on the validation set for each λ
4. Choose the λ that minimizes the validation error

This process helps balance the bias-variance tradeoff: too small λ may lead to overfitting, while too large λ may cause underfitting.

3 Bayes' theorem

3.1 Calculating $p(r)$

We can calculate $p(r)$ using the law of total probability:

$$p(r) = p(r|\leftarrow)p(\leftarrow) + p(r|\rightarrow)p(\rightarrow) \quad (29)$$

Given that $p(r|s)$ follows a Gaussian distribution for each stimulus s :

$$p(r|s) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(r-\mu_s)^2}{2\sigma^2}\right) \quad (30)$$

Therefore:

$$p(r) = p(\leftarrow) \cdot \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(r-\mu_{\leftarrow})^2}{2\sigma^2}\right) + p(\rightarrow) \cdot \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(r-\mu_{\rightarrow})^2}{2\sigma^2}\right) \quad (31)$$

$$= \frac{1}{\sqrt{2\pi}\sigma} \left[p(\leftarrow) \exp\left(-\frac{(r-\mu_{\leftarrow})^2}{2\sigma^2}\right) + p(\rightarrow) \exp\left(-\frac{(r-\mu_{\rightarrow})^2}{2\sigma^2}\right) \right] \quad (32)$$

This is a mixture of two Gaussians with means μ_{\leftarrow} and μ_{\rightarrow} , standard deviation σ , and mixing proportions $p(\leftarrow)$ and $p(\rightarrow)$.

Let the standardized separation be

$$d = \frac{|\mu_{\rightarrow} - \mu_{\leftarrow}|}{\sigma}.$$

When $d \lesssim 1$ the two bells overlap so much that the mixture shows just one, slightly broader peak (leaning toward the heavier-weighted side if the mixing proportions differ). For $1 \lesssim d \lesssim 2$ this peak develops a flat top or shoulder. Once $d \gtrsim 2$ (and even larger if one component is much rarer) the curve splits into two distinct peaks; the taller peak lies under the component with the larger mixing weight.

3.2 Posterior calculation with equal priors

With $p(\leftarrow) = p(\rightarrow) = 1/2$, the posterior is:

$$p(\leftarrow|r) = \frac{p(r|\leftarrow)p(\leftarrow)}{p(r)} \quad (33)$$

$$= \frac{p(r|\leftarrow) \cdot \frac{1}{2}}{p(r|\leftarrow) \cdot \frac{1}{2} + p(r|\rightarrow) \cdot \frac{1}{2}} \quad (34)$$

$$= \frac{p(r|\leftarrow)}{p(r|\leftarrow) + p(r|\rightarrow)} \quad (35)$$

$$= \frac{\exp\left(-\frac{(r-\mu_{\leftarrow})^2}{2\sigma^2}\right)}{\exp\left(-\frac{(r-\mu_{\leftarrow})^2}{2\sigma^2}\right) + \exp\left(-\frac{(r-\mu_{\rightarrow})^2}{2\sigma^2}\right)} \quad (36)$$

$$= \frac{1}{1 + \exp\left(-\frac{(r-\mu_{\leftarrow})^2 - (r-\mu_{\rightarrow})^2}{2\sigma^2}\right)} \quad (37)$$

$$= \frac{1}{1 + \exp\left(-\frac{2r(\mu_{\rightarrow} - \mu_{\leftarrow}) + \mu_{\leftarrow}^2 - \mu_{\rightarrow}^2}{2\sigma^2}\right)} \quad (38)$$

Similarly, for $p(\rightarrow|r)$ we get:

$$p(\rightarrow|r) = \frac{1}{1 + \exp\left(-\frac{2r(\mu_{\leftarrow} - \mu_{\rightarrow}) + \mu_{\rightarrow}^2 - \mu_{\leftarrow}^2}{2\sigma^2}\right)} \quad (39)$$

The posterior would produce a sigmoid-shaped curve, crossing at the midpoint between μ_{\leftarrow} and μ_{\rightarrow} , where $p(\leftarrow|r) = p(\rightarrow|r) = 0.5$.

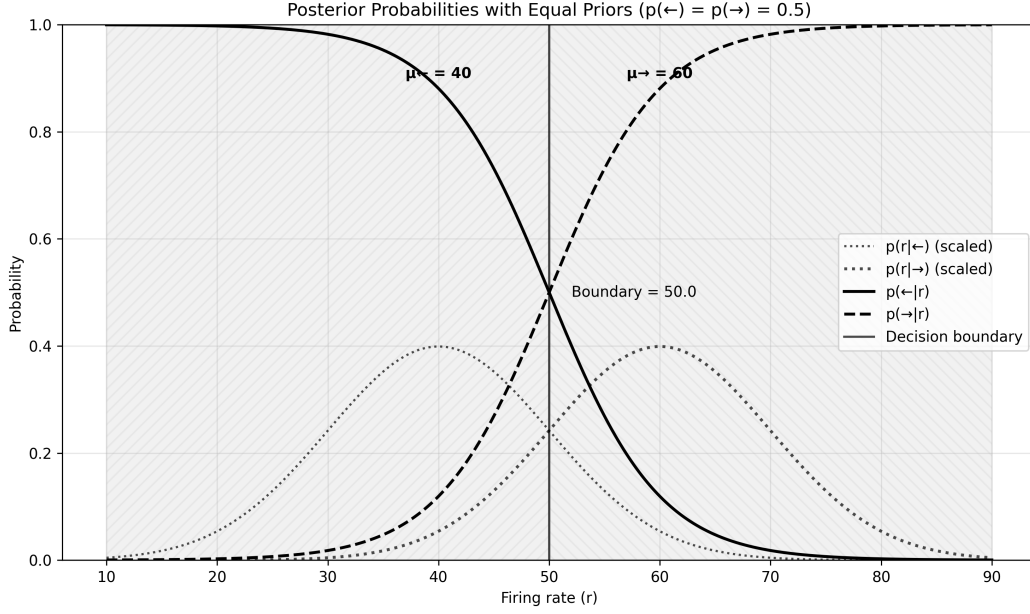


Figure 1: Posterior probabilities for the equal-prior case, $p(\leftarrow) = p(\rightarrow) = 0.5$. The solid black curve shows $p(\leftarrow | r)$ and the dashed black curve shows $p(\rightarrow | r)$. Grey dotted lines depict the (rescaled) likelihoods $p(r | \leftarrow)$ and $p(r | \rightarrow)$. The two posteriors are complementary sigmoids that intersect at $r = 50$, the midpoint between the means $\mu_{\leftarrow} = 40$ and $\mu_{\rightarrow} = 60$. The vertical line marks this optimal decision boundary.

3.3 Effect of changing the prior

When the priors are unequal, the posterior probabilities shift accordingly:

- When $p(\leftarrow) > p(\rightarrow)$, the sigmoid curve shifts to the right, making leftward motion more likely for a wider range of firing rates
- When $p(\rightarrow) > p(\leftarrow)$, the curve shifts to the left, favoring rightward motion for more firing rates

The crossing point of the posteriors (where $p(\leftarrow | r) = p(\rightarrow | r)$) moves away from the midpoint between the means.

3.4 Decision rule analysis

The decision rule $r > \frac{1}{2}(\mu_{\leftarrow} + \mu_{\rightarrow})$ corresponds to choosing the stimulus with the higher likelihood when priors are equal. This is equivalent to the point where the posteriors cross ($p(\leftarrow | r) = p(\rightarrow | r) = 0.5$) under equal priors.

When prior probabilities are unequal, the fixed decision threshold no longer aligns with the point where the posterior probabilities intersect. As a result, the decision rule ignores prior information about stimulus frequency, potentially leading to suboptimal performance and increased error rates.

By using a simple threshold instead of the full posterior, we lose the ability to properly account for prior probabilities, the ability to express confidence in our decisions, the probabilistic information provided by the posterior and the flexibility to optimize different decision criteria (e.g., minimizing total error vs. avoiding false alarms)

4 Linear discriminant analysis

4.1 Computing the log-likelihood ratio

The log-likelihood ratio is:

$$l(r) = \log \frac{p(r | \leftarrow)}{p(r | \rightarrow)} \quad (40)$$

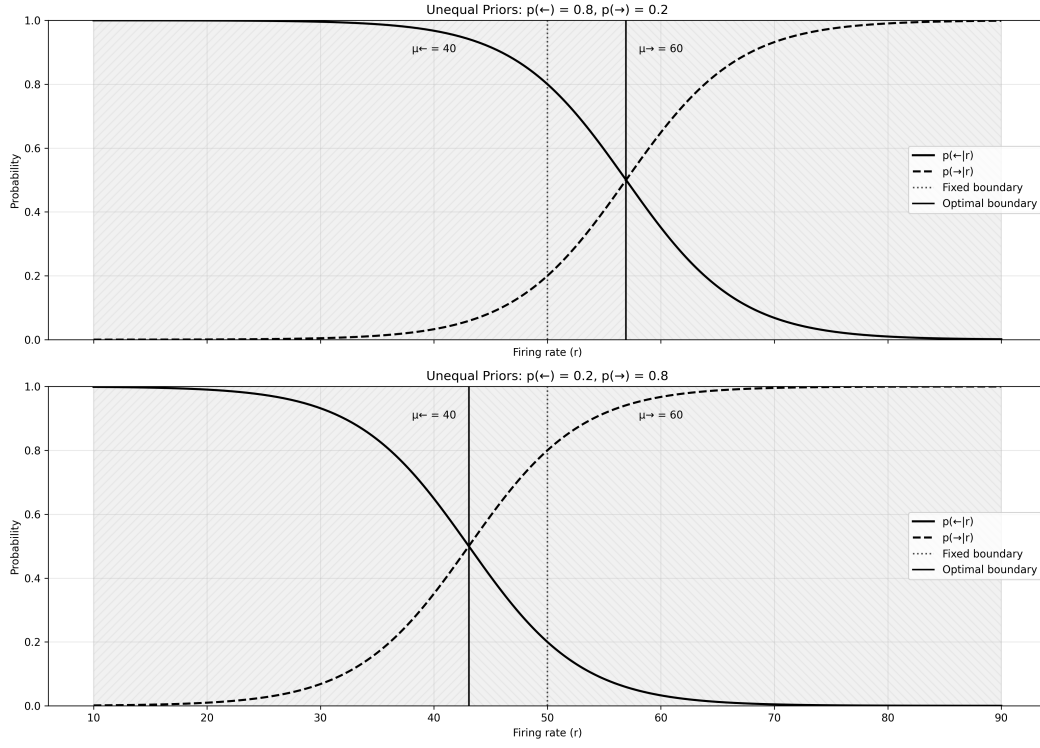


Figure 2: Posterior probabilities with unequal priors. Top: When $p(\leftarrow) = 0.8, p(\rightarrow) = 0.2$, the decision boundary shifts to the right, favoring leftward motion decisions for a wider range of firing rates. Bottom: When $p(\leftarrow) = 0.2, p(\rightarrow) = 0.8$, the decision boundary shifts to the left, favoring rightward motion decisions.

Substituting the multivariate Gaussian distributions:

$$l(r) = \log \frac{\frac{1}{(2\pi)^{N/2}\sqrt{\det C}} \exp\left(-\frac{1}{2}(r - \mu_{\leftarrow})^T C^{-1}(r - \mu_{\leftarrow})\right)}{\frac{1}{(2\pi)^{N/2}\sqrt{\det C}} \exp\left(-\frac{1}{2}(r - \mu_{\rightarrow})^T C^{-1}(r - \mu_{\rightarrow})\right)} \quad (41)$$

$$= \log \exp\left(-\frac{1}{2}(r - \mu_{\leftarrow})^T C^{-1}(r - \mu_{\leftarrow}) + \frac{1}{2}(r - \mu_{\rightarrow})^T C^{-1}(r - \mu_{\rightarrow})\right) \quad (42)$$

$$= -\frac{1}{2}(r - \mu_{\leftarrow})^T C^{-1}(r - \mu_{\leftarrow}) + \frac{1}{2}(r - \mu_{\rightarrow})^T C^{-1}(r - \mu_{\rightarrow}) \quad (43)$$

Expanding the quadratic forms:

$$l(r) = -\frac{1}{2} [r^T C^{-1} r - r^T C^{-1} \mu_{\leftarrow} - \mu_{\leftarrow}^T C^{-1} r + \mu_{\leftarrow}^T C^{-1} \mu_{\leftarrow}] \quad (44)$$

$$+ \frac{1}{2} [r^T C^{-1} r - r^T C^{-1} \mu_{\rightarrow} - \mu_{\rightarrow}^T C^{-1} r + \mu_{\rightarrow}^T C^{-1} \mu_{\rightarrow}] \quad (45)$$

The $r^T C^{-1} r$ terms cancel out, and since C is symmetric, $r^T C^{-1} \mu = \mu^T C^{-1} r$:

$$l(r) = \frac{1}{2} [2r^T C^{-1}(\mu_{\leftarrow} - \mu_{\rightarrow}) - \mu_{\leftarrow}^T C^{-1} \mu_{\leftarrow} + \mu_{\rightarrow}^T C^{-1} \mu_{\rightarrow}] \quad (46)$$

$$= r^T C^{-1}(\mu_{\leftarrow} - \mu_{\rightarrow}) - \frac{1}{2} [\mu_{\leftarrow}^T C^{-1} \mu_{\leftarrow} - \mu_{\rightarrow}^T C^{-1} \mu_{\rightarrow}] = 0 \quad (47)$$

This is a linear equation in r , which can be written as:

$$r^T w + b = 0 \quad (48)$$

where:

$$w = C^{-1}(\mu_{\leftarrow} - \mu_{\rightarrow}) \quad (49)$$

$$b = -\frac{1}{2} [\mu_{\leftarrow}^T C^{-1} \mu_{\leftarrow} - \mu_{\rightarrow}^T C^{-1} \mu_{\rightarrow}] \quad (50)$$

This defines a hyperplane (linear) in the N -dimensional firing rate space. This hyperplane separates the regions where $l(r) > 0$ (classified as leftward motion) from regions where $l(r) < 0$ (classified as rightward motion).

4.2 Decision boundary with uncorrelated neurons

With $p(\leftarrow) = p(\rightarrow) = 1/2$ and a diagonal covariance matrix $C = \text{diag}(\sigma_1^2, \sigma_2^2)$, we have:

$$C^{-1} = \text{diag}\left(\frac{1}{\sigma_1^2}, \frac{1}{\sigma_2^2}\right) \quad (51)$$

Let's denote $\mu_{\leftarrow} = [\mu_{\leftarrow,1}, \mu_{\leftarrow,2}]^T$ and $\mu_{\rightarrow} = [\mu_{\rightarrow,1}, \mu_{\rightarrow,2}]^T$.

The decision boundary equation becomes:

$$\begin{bmatrix} r_1 & r_2 \end{bmatrix} \begin{bmatrix} \frac{1}{\sigma_1^2} & 0 \\ 0 & \frac{1}{\sigma_2^2} \end{bmatrix} \begin{bmatrix} \mu_{\leftarrow,1} - \mu_{\rightarrow,1} \\ \mu_{\leftarrow,2} - \mu_{\rightarrow,2} \end{bmatrix} \quad (52)$$

$$- \frac{1}{2} \begin{bmatrix} \mu_{\leftarrow,1} & \mu_{\leftarrow,2} \end{bmatrix} \begin{bmatrix} \frac{1}{\sigma_1^2} & 0 \\ 0 & \frac{1}{\sigma_2^2} \end{bmatrix} \begin{bmatrix} \mu_{\leftarrow,1} \\ \mu_{\leftarrow,2} \end{bmatrix} - \begin{bmatrix} \mu_{\rightarrow,1} & \mu_{\rightarrow,2} \end{bmatrix} \begin{bmatrix} \frac{1}{\sigma_1^2} & 0 \\ 0 & \frac{1}{\sigma_2^2} \end{bmatrix} \begin{bmatrix} \mu_{\rightarrow,1} \\ \mu_{\rightarrow,2} \end{bmatrix} = 0 \quad (53)$$

Simplifying:

$$\frac{r_1}{\sigma_1^2}(\mu_{\leftarrow,1} - \mu_{\rightarrow,1}) + \frac{r_2}{\sigma_2^2}(\mu_{\leftarrow,2} - \mu_{\rightarrow,2}) \quad (54)$$

$$- \frac{1}{2} \left[\frac{\mu_{\leftarrow,1}^2}{\sigma_1^2} + \frac{\mu_{\leftarrow,2}^2}{\sigma_2^2} - \frac{\mu_{\rightarrow,1}^2}{\sigma_1^2} - \frac{\mu_{\rightarrow,2}^2}{\sigma_2^2} \right] = 0 \quad (55)$$

This is a straight line in the (r_1, r_2) plane, which can be written in the form:

$$ar_1 + br_2 + c = 0 \quad (56)$$

where:

$$a = \frac{\mu_{\leftarrow,1} - \mu_{\rightarrow,1}}{\sigma_1^2} \quad (57)$$

$$b = \frac{\mu_{\leftarrow,2} - \mu_{\rightarrow,2}}{\sigma_2^2} \quad (58)$$

$$c = -\frac{1}{2} \left[\frac{\mu_{\leftarrow,1}^2 - \mu_{\rightarrow,1}^2}{\sigma_1^2} + \frac{\mu_{\leftarrow,2}^2 - \mu_{\rightarrow,2}^2}{\sigma_2^2} \right] \quad (59)$$

The slope of this line is $-\frac{a}{b} = -\frac{\sigma_2^2(\mu_{\leftarrow,1} - \mu_{\rightarrow,1})}{\sigma_1^2(\mu_{\leftarrow,2} - \mu_{\rightarrow,2})}$.

This boundary optimally separates the two Gaussian distributions in the firing rate space. The orientation of the line depends on the difference in means and the variances of the two neurons (Fig 3).

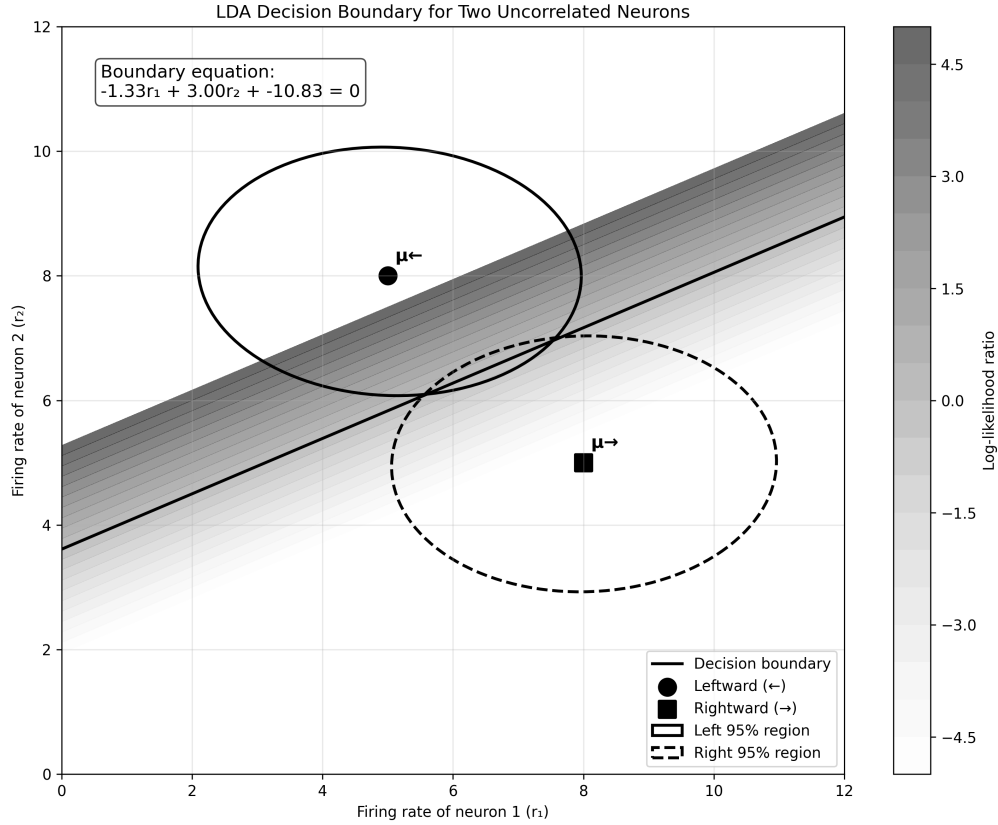


Figure 3: This figure illustrates the LDA decision boundary for a 2D case with two uncorrelated neurons responding to different motion directions. The grayscale contour plot represents the log-likelihood ratio between leftward and rightward motion. The parameters are: mean firing rates for leftward motion $\mu_{\leftarrow} = [5, 8]$, mean firing rates for rightward motion $\mu_{\rightarrow} = [8, 5]$, with standard deviations $\sigma_1 = 1.5$ and $\sigma_2 = 1.0$. The solid black line represents the decision boundary, which optimally separates the two motion categories. The 95% confidence regions for each motion direction are shown as ellipses. The decision boundary is defined by a linear equation based on the firing rates of both neurons, demonstrating how the neurons could optimally decode motion direction from neural activity patterns.