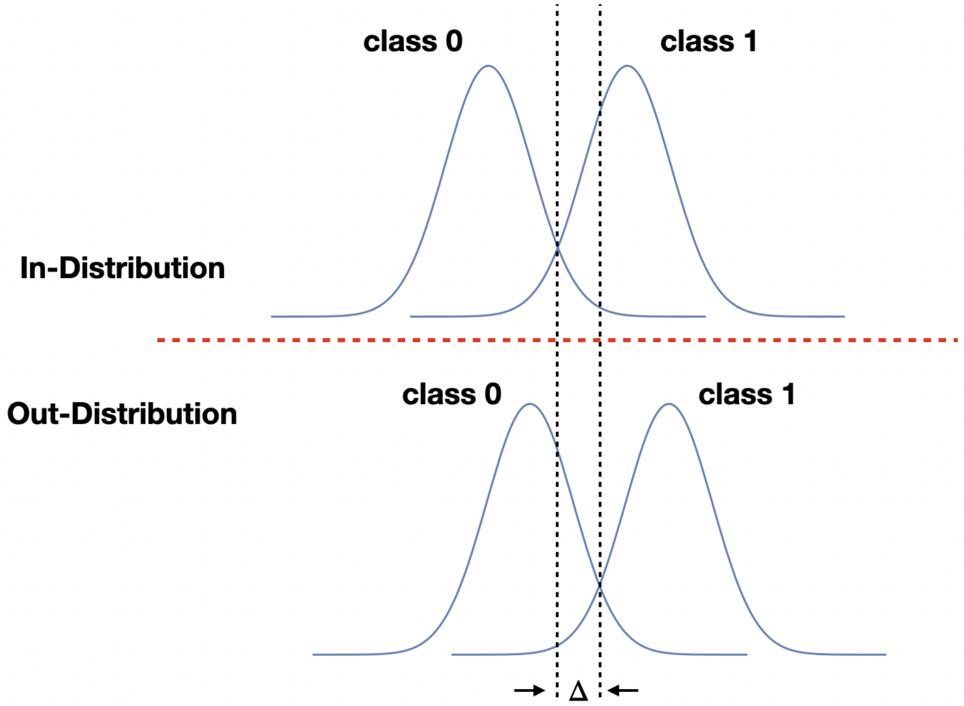
# Out-of-Distribution Learning

# Gaussian Tasks Experiment

- Consider an in-distribution task that consists of two class conditional gaussians.
- Now, consider an out-of-distribution task similar to the above task, but whose center is displaced by an amount $\Delta$.
- The amount $\Delta$ reflects the "similarity" between the two tasks.



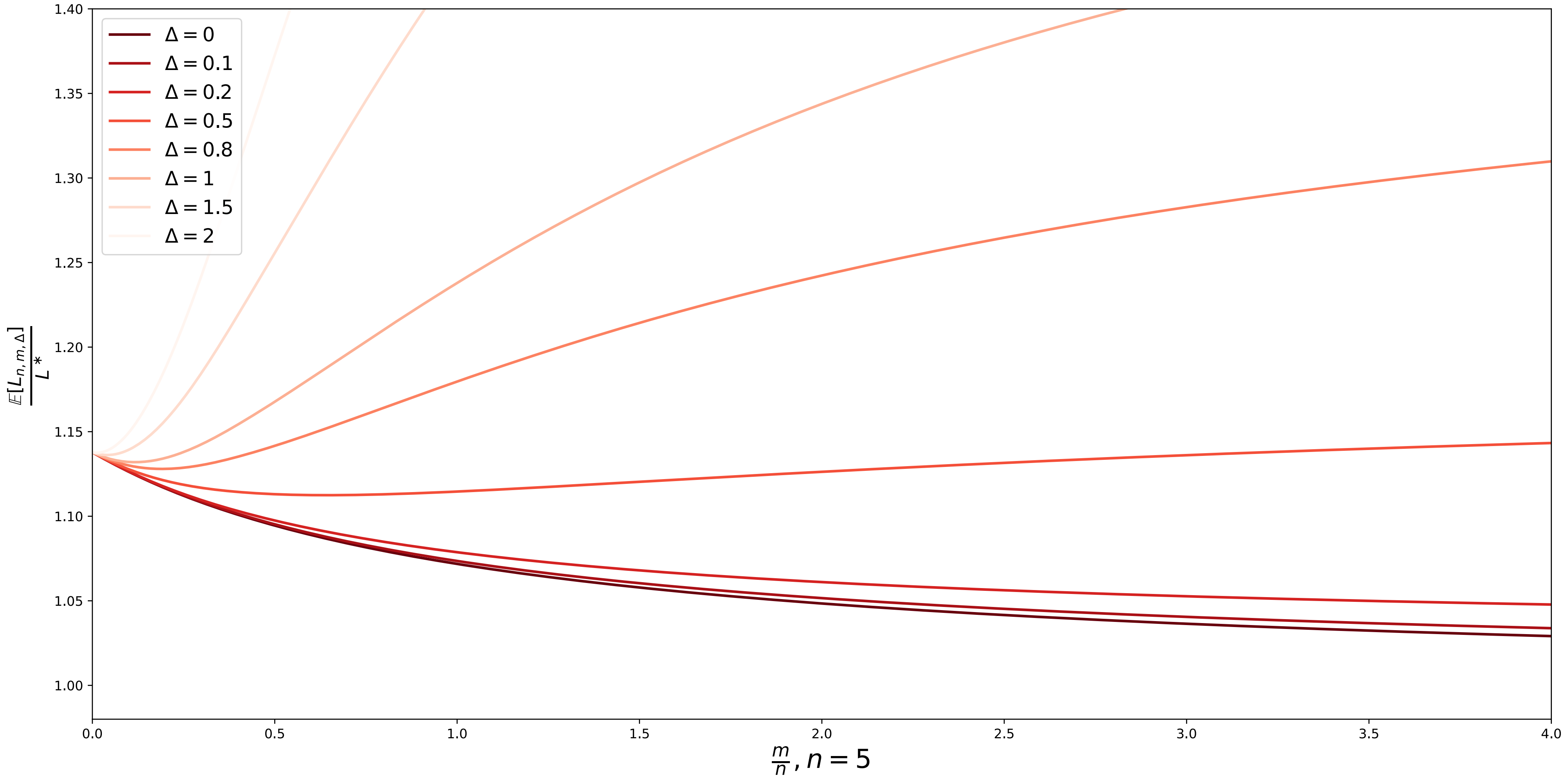
# Gaussian Tasks Experiment

- We have access to $n$ samples from the in-distribution task, and $m$ samples from the out-of-distribution task.
- Using both the in-distribution and out-of-distribution samples, we train a classifier $h$ aimed at the in-distribution classification task.
- Let's denote the classification error of $h$ by $\mathbb{E}[L_{n,m,\Delta}]$.

## Gaussian Tasks Experiment

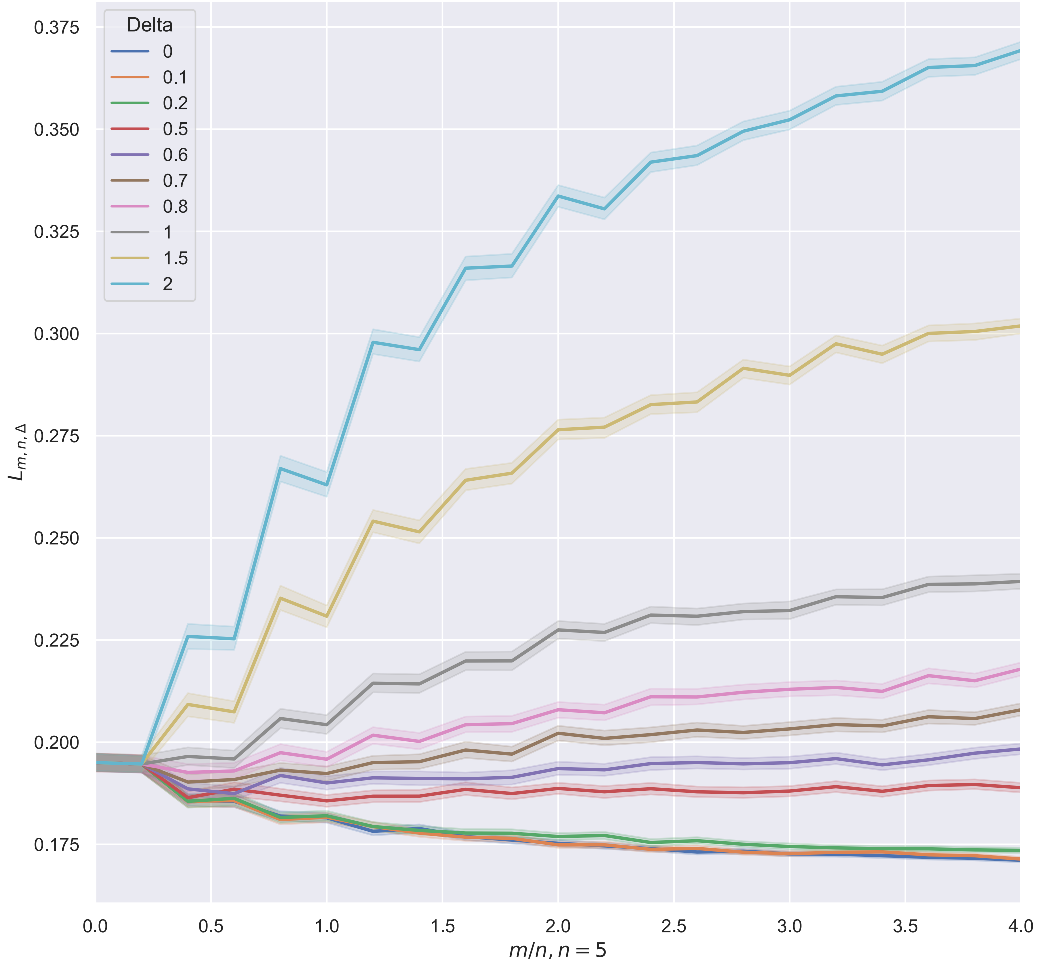
- Let $n$ be a small fixed constant. We hypothesize that,
    - For very small $\Delta$, as we add more out-of-distribution data (as $m$ increases) the $\mathbb{E}[L_{n,m,\Delta}]$ would decrease.
    - For moderately large $\Delta$, as we add more out-of-distribution data (as $m$ increases) the $\mathbb{E}[L_{n,m,\Delta}]$ would initially decrease and start increasing later. The initial decrease is due to the reduction in the variance of $h$. The later increase is due to the increase in bias of $h$ caused by the out-of-distribution samples.
    - For very large $\Delta$, as we add more out-of-distribution data (as $m$ increases) the $\mathbb{E}[L_{n,m,\Delta}]$ would keep increasing.

# Gaussian Tasks Experiment



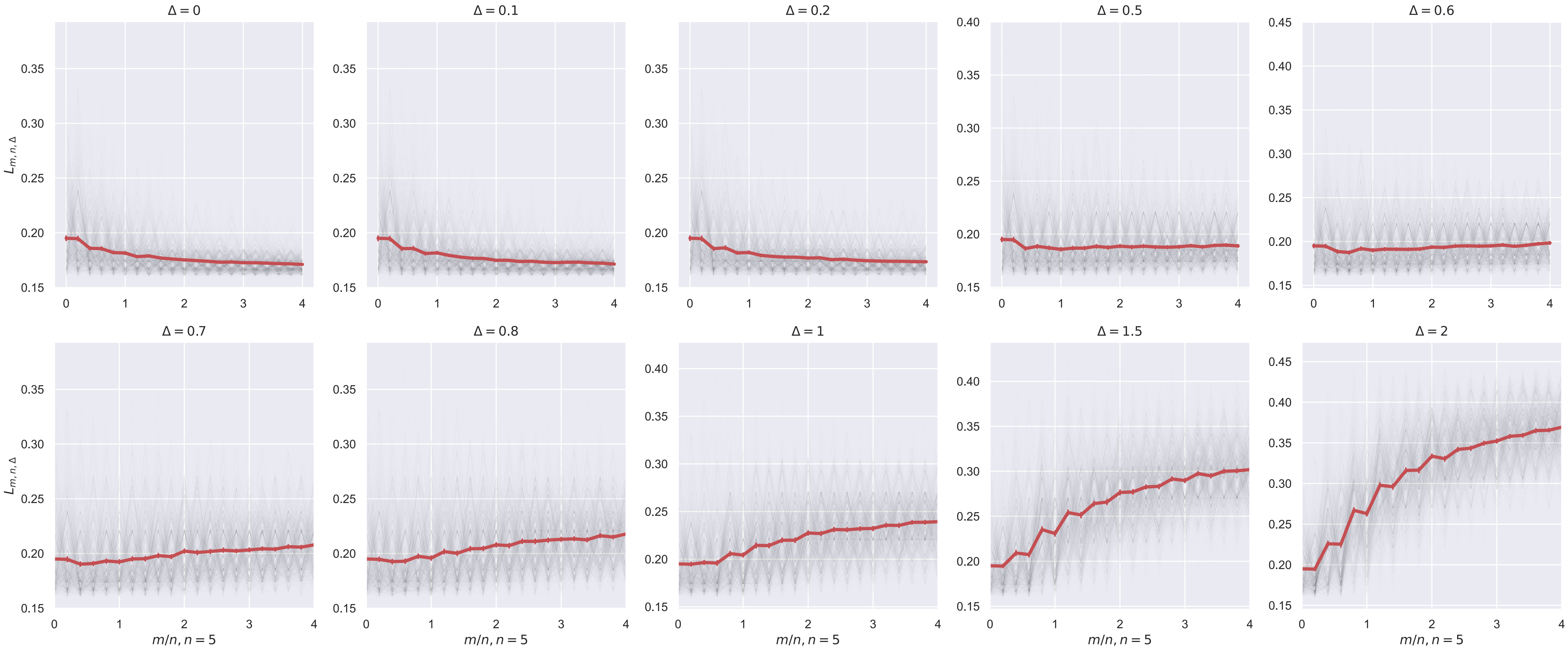
# Gaussian Tasks Experiment

- Number of replicates: 1000



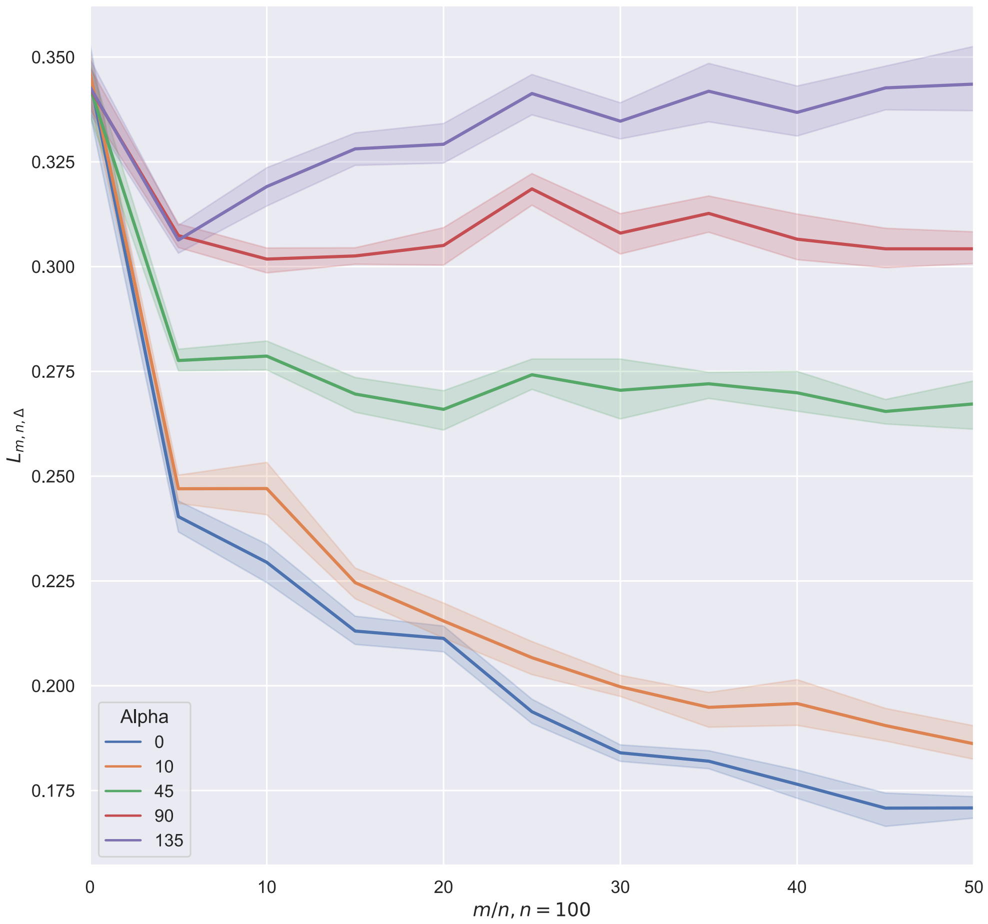
# Gaussian Tasks Experiment

- Number of replicates: 1000



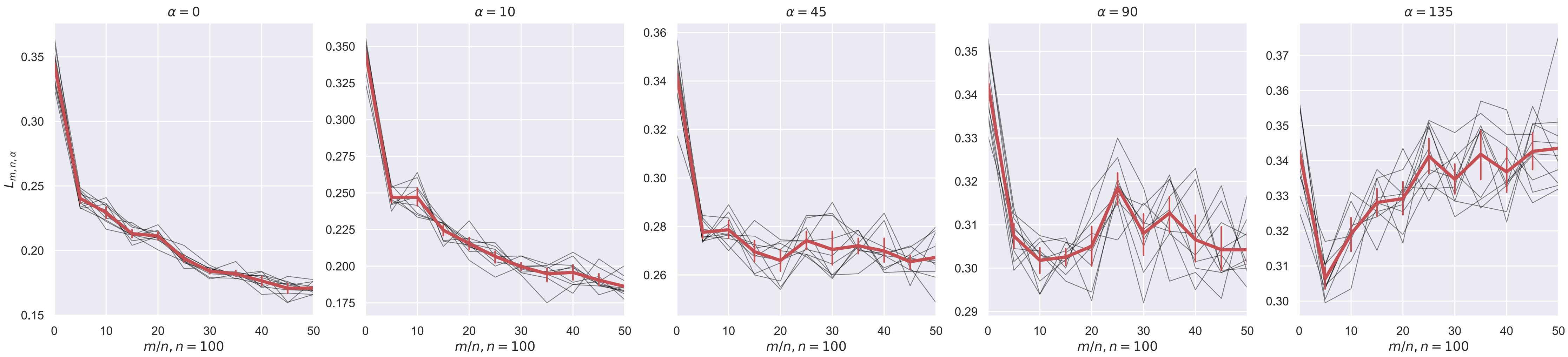
# Bird vs. Cat & $\alpha$-Rotated Bird vs. Cat (Single-Head Network)

- Number of replicates: 10, Network: SmallConv



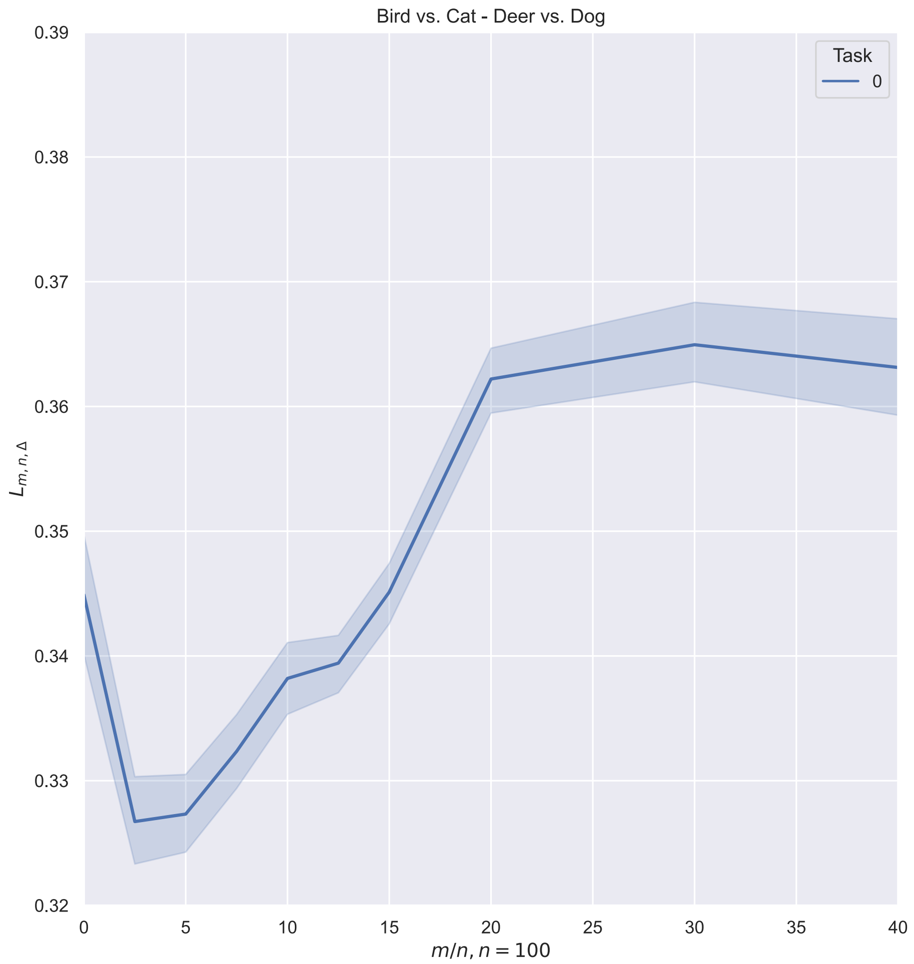
# Bird vs. Cat & $\alpha$-Rotated Bird vs. Cat (Single-Head Network)

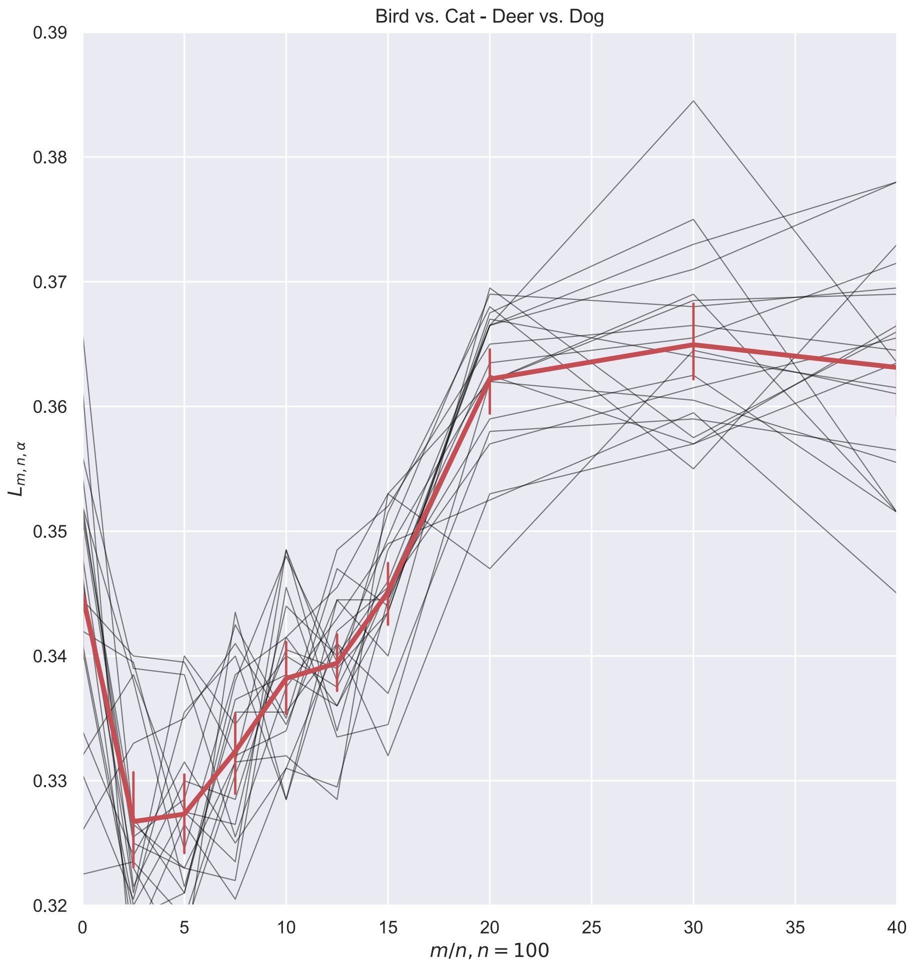
- Number of replicates: 10, Network: SmallConv



# Task 2: Bird vs. Cat & Task 3: Deer vs. Dog (Single-Head Network)

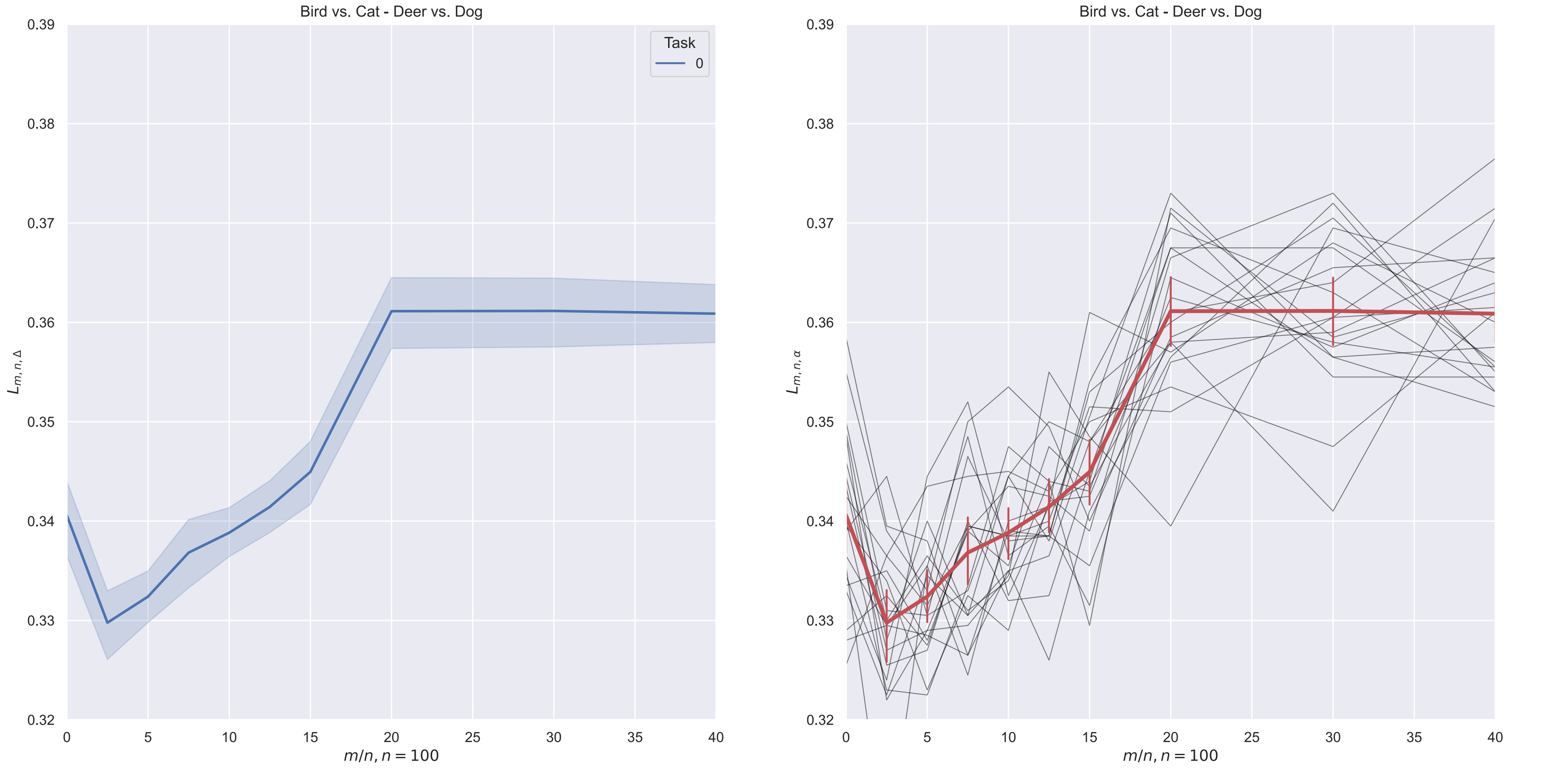
- Number of replicates: 20, Network: SmallConv



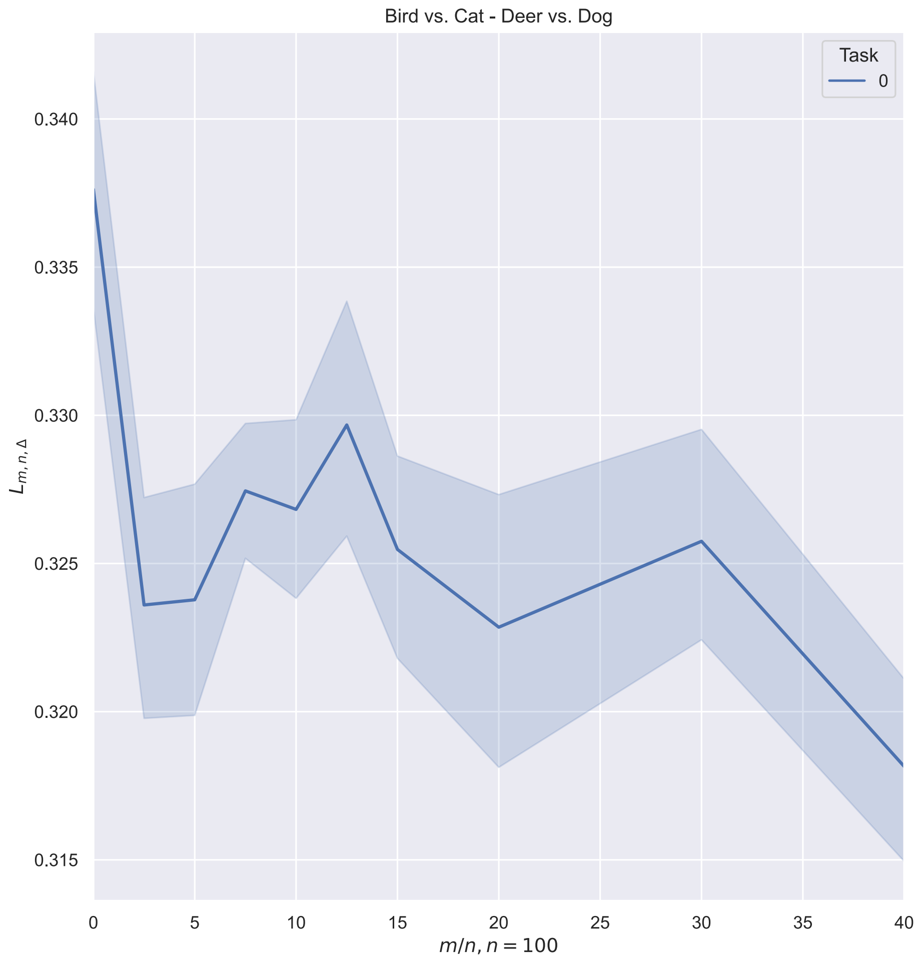
# Task 2: Bird vs. Cat & Task 3: Deer vs. Dog (Single-Head Network)

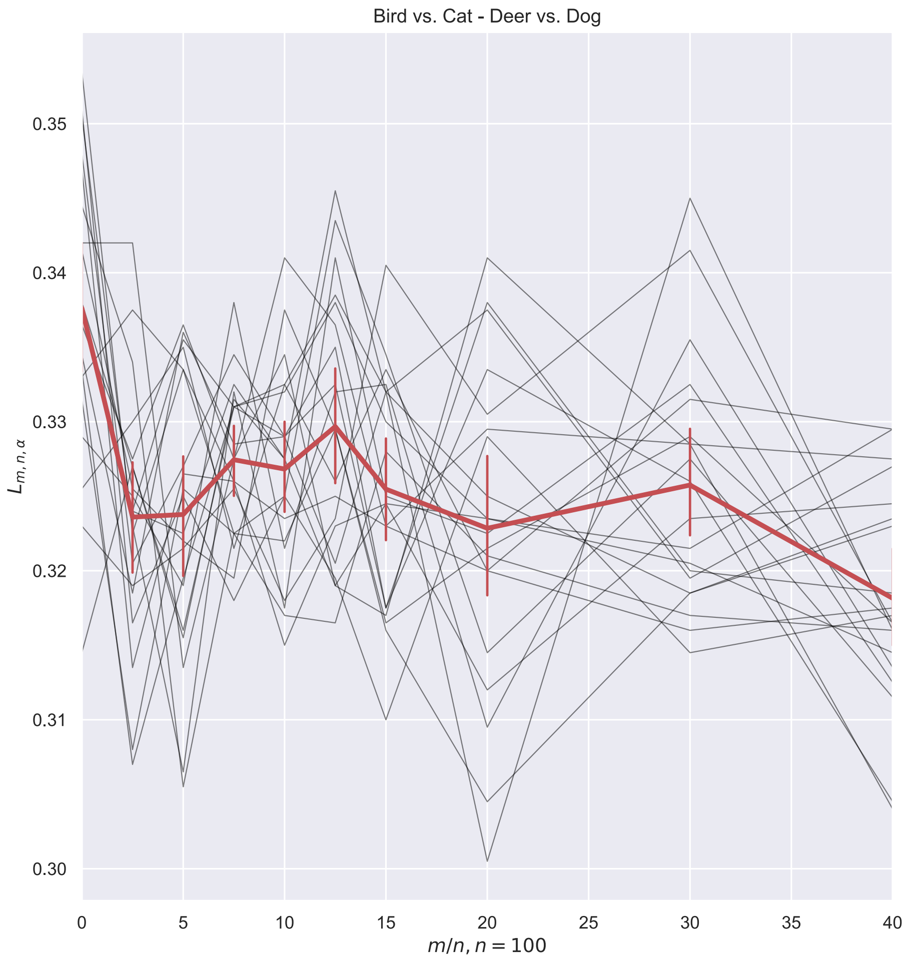
- Number of replicates: 20, Network: SmallConv, each model was trained for 100 epochs



# Task 2: Bird vs. Cat & Task 3: Deer vs. Dog (Multi-Head Network)

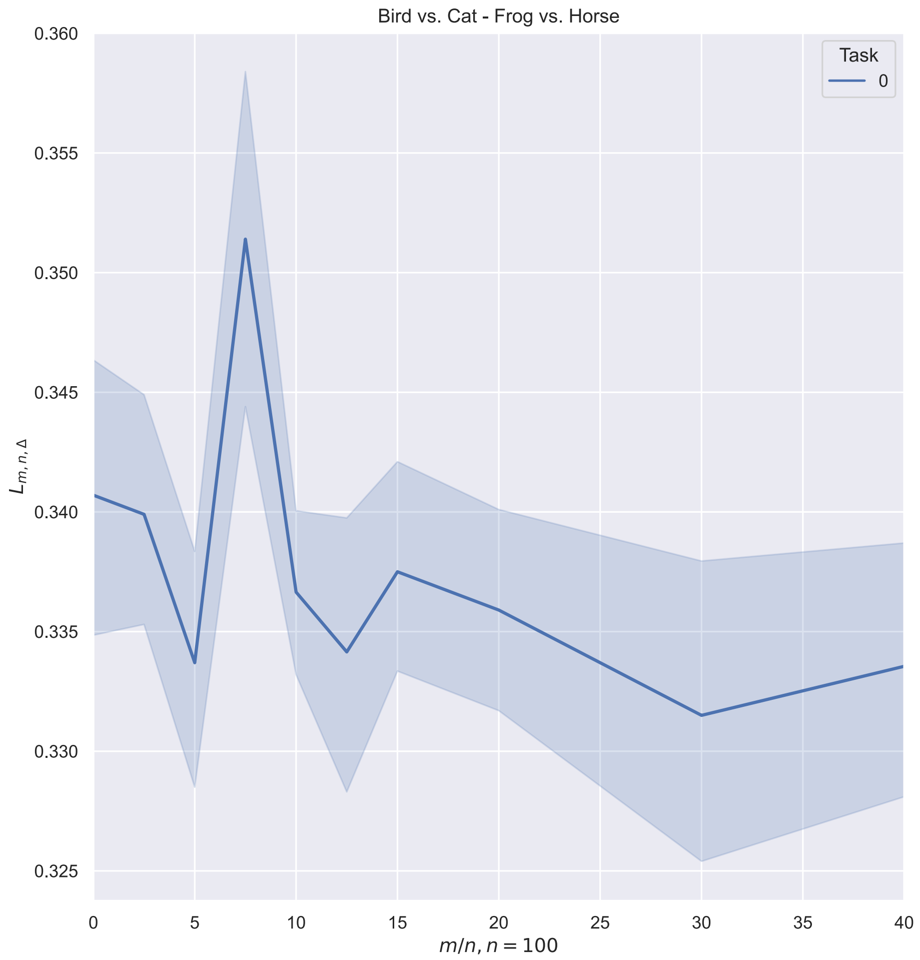
- Number of replicates: 20, Network: SmallConv

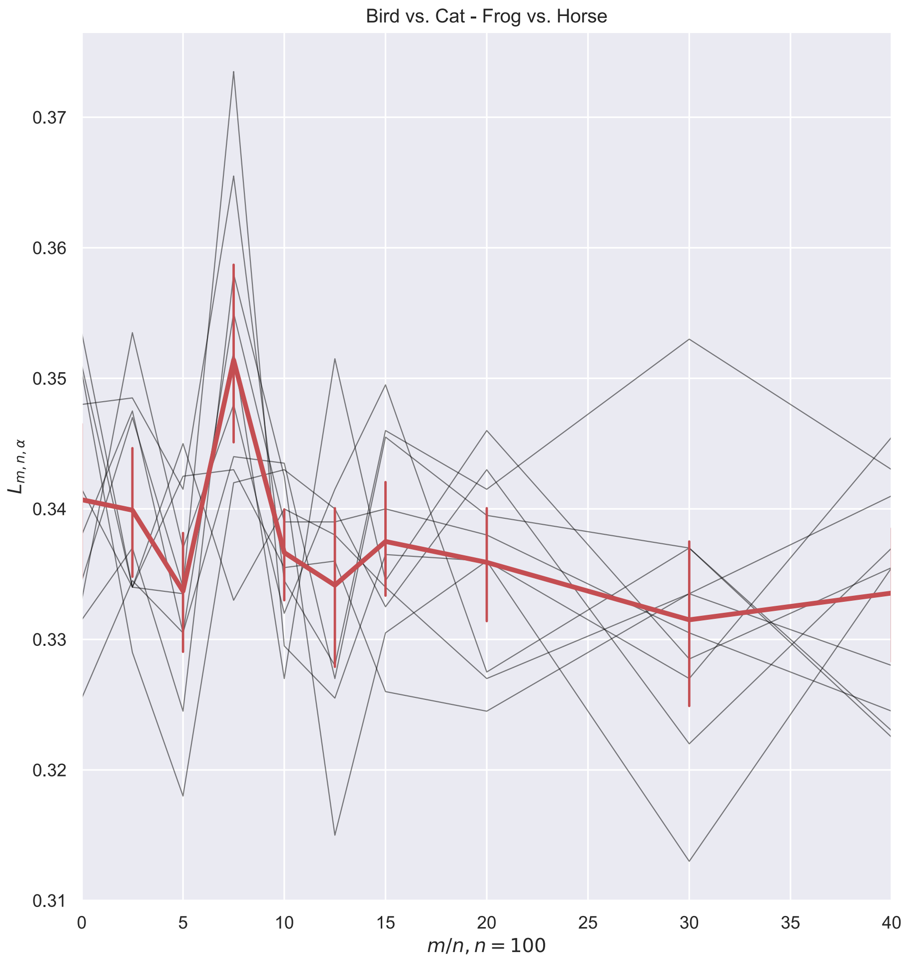


# Task 2: Bird vs. Cat & Task 4: Frog vs. Horse (Multi-Head Network)

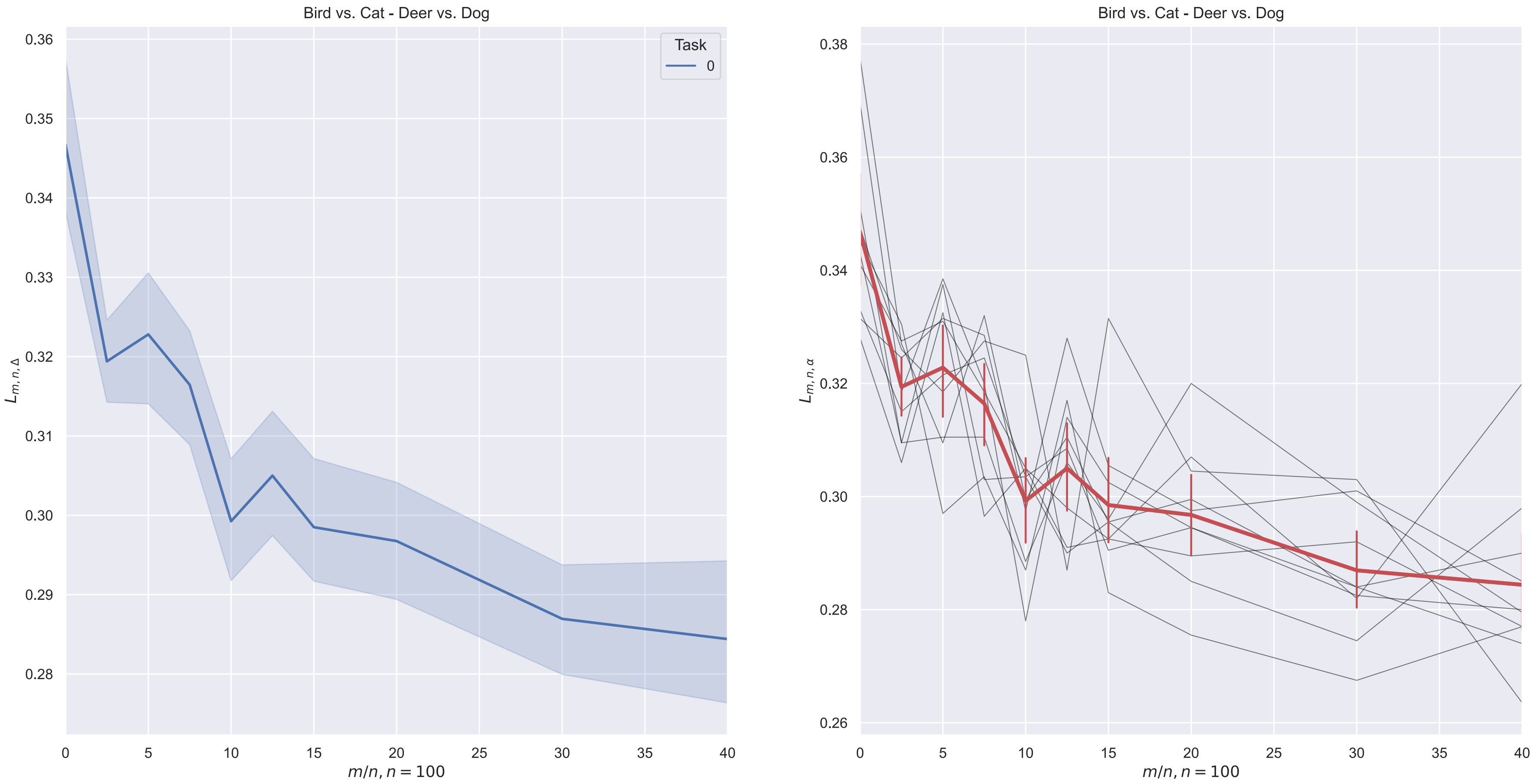
- Number of replicates: 20, Network: SmallConv



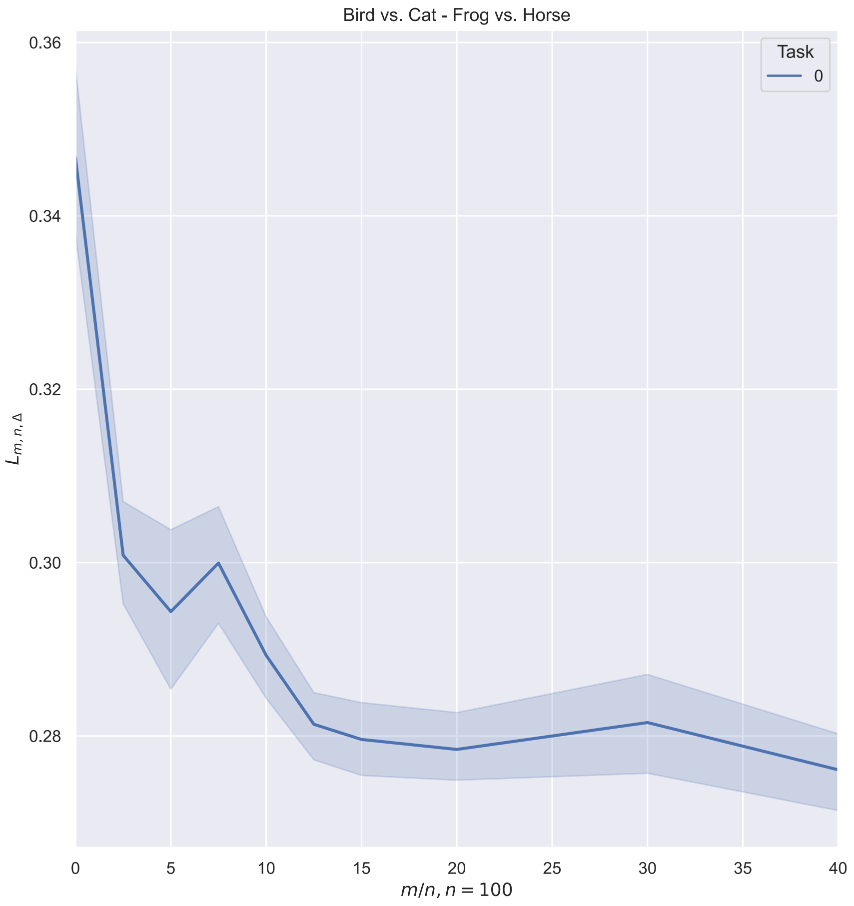
# Task 2: Bird vs. Cat & Task 3: Deer vs. Dog (Multi-Head Network)

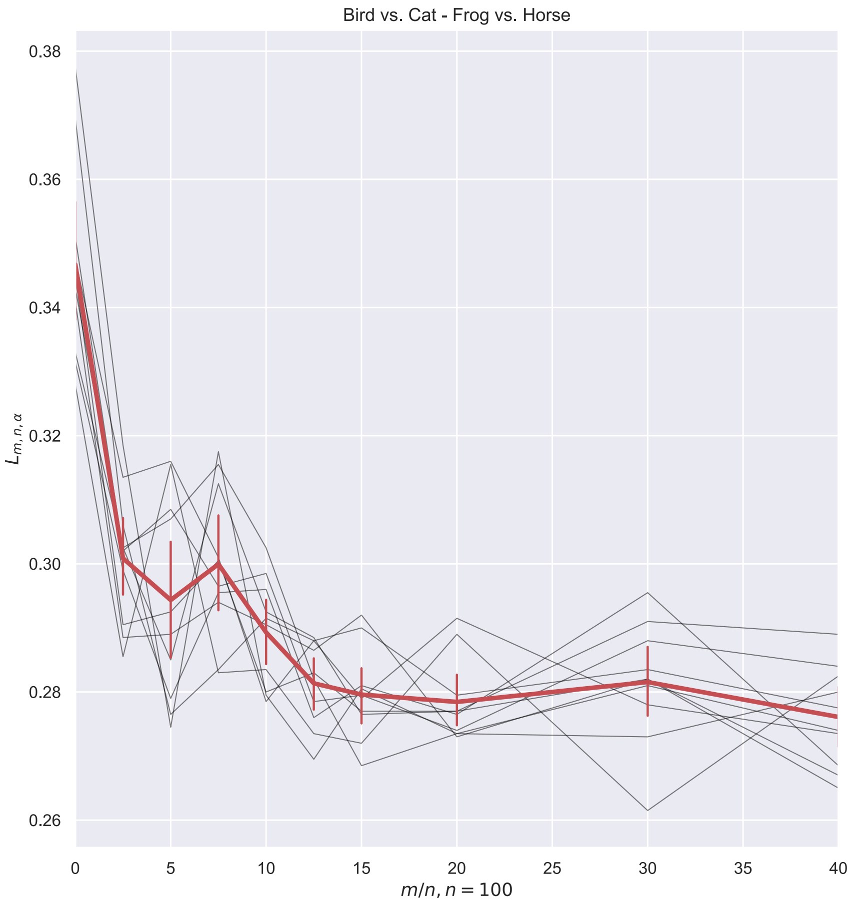
- Number of replicates: 10, Network: Wide Res-Net



# Task 2: Bird vs. Cat & Task 4: Frog vs. Horse (Multi-Head Network)

- Number of replicates: 10, Network: Wide Res-Net

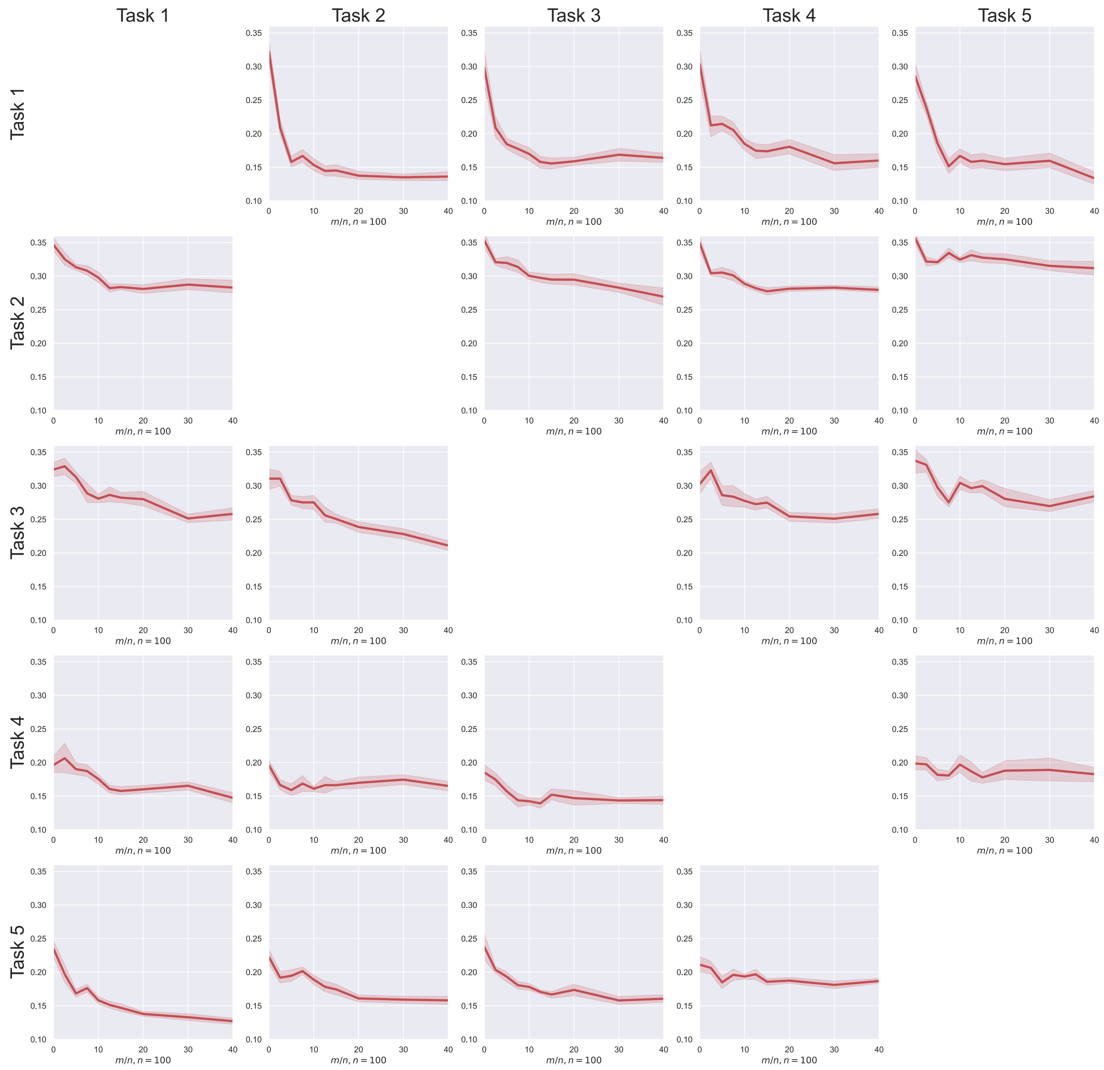


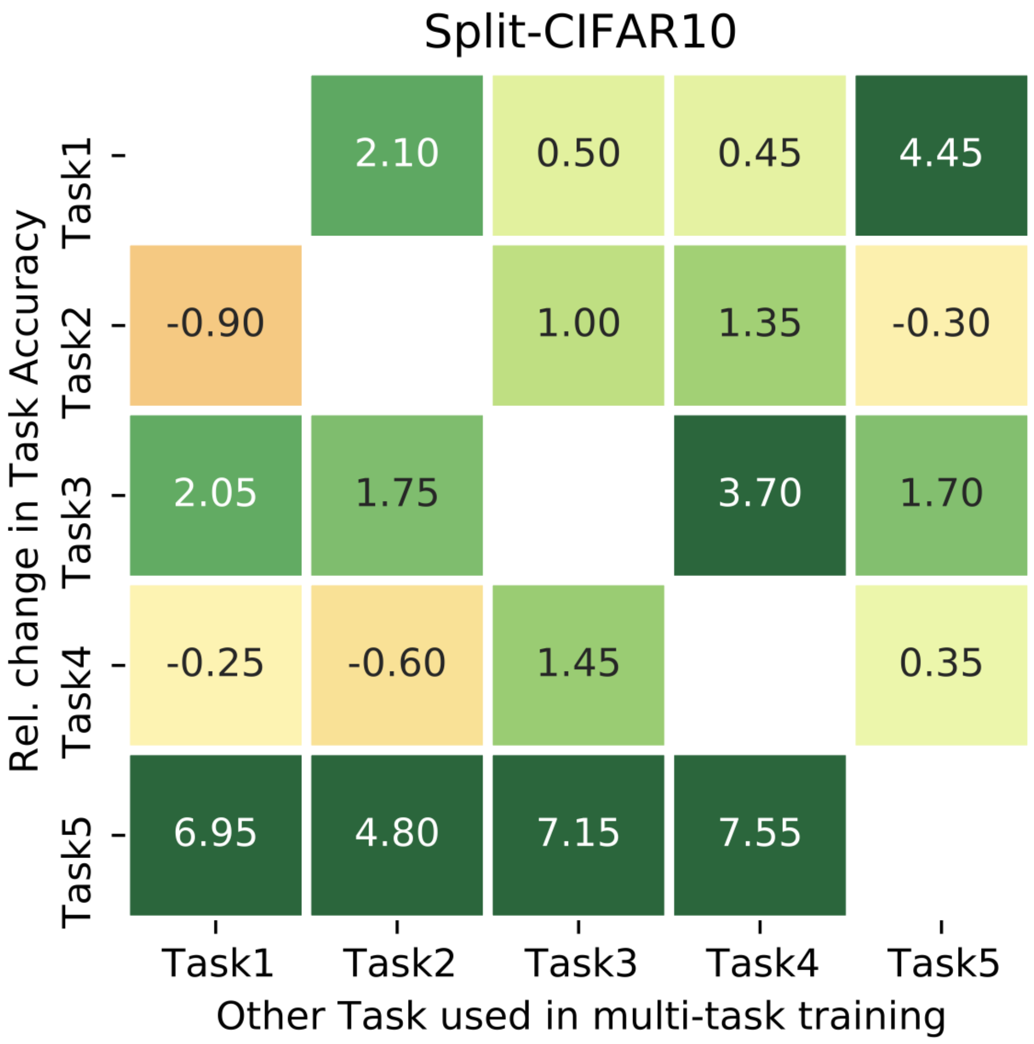
# Task 4: Frog vs. Horse & Task 2: Bird vs. Cat (Multi-Head Network)

- Number of replicates: 10, Network: Wide Res-Net

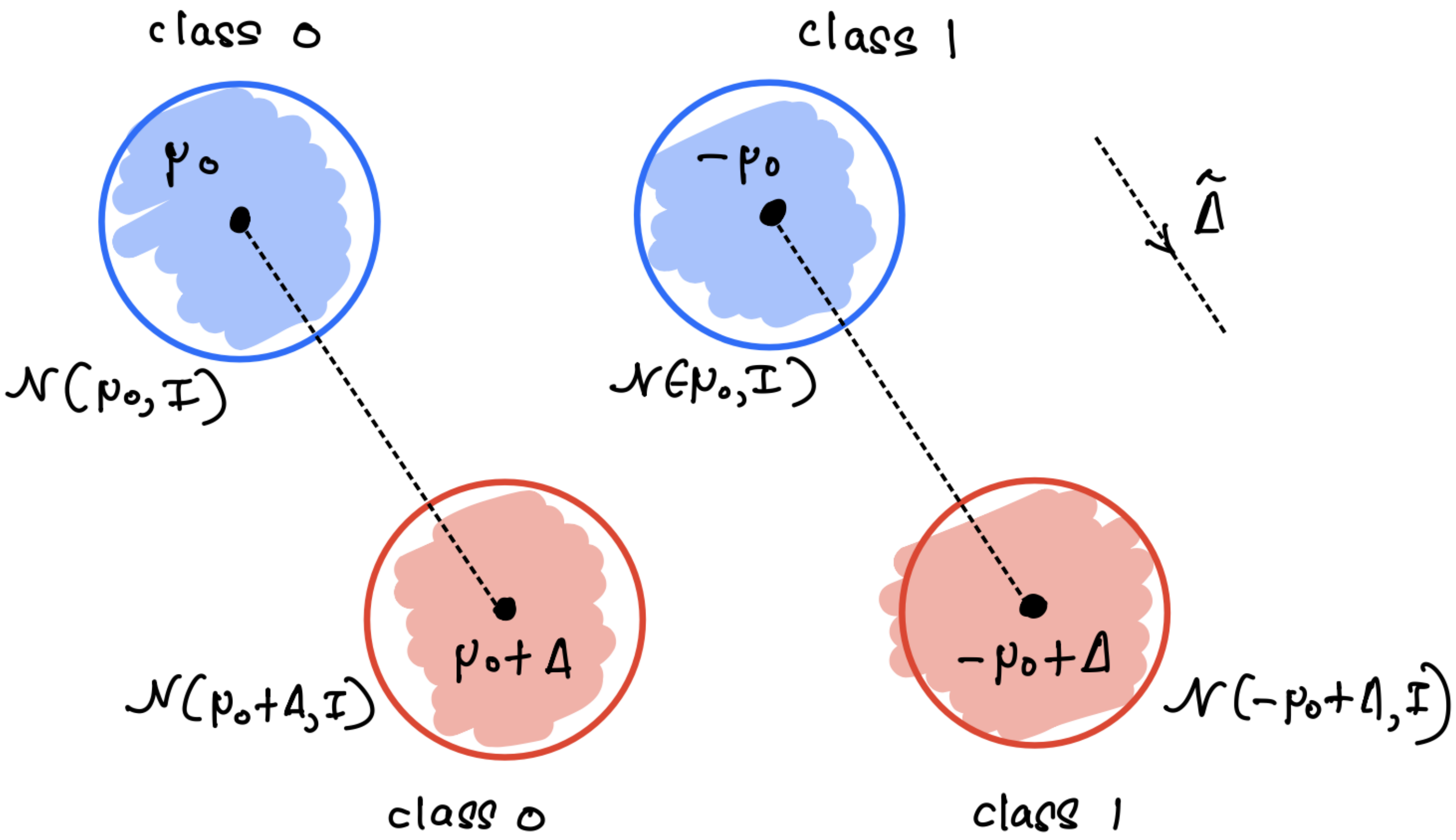
# CIFAR-10 Tasks (Multi-Head Network)



# CIFAR-10 Tasks (Multi-Head Network)



# Bivariate LDA



# Bivariate LDA

- In-distribution samples (class 0): $X_1, \dots, X_{n/2} \sim \mathcal{N}(\mu, I)$
- In-distribution samples (class 1): $X_{n/2+1}, \dots, X_n \sim \mathcal{N}(-\mu, I)$
- Out-of-distribution samples (class 0): $X_{n+1}, \dots, X_{n+m/2} \sim \mathcal{N}(\mu + \Delta, I)$
- In-distribution samples (class 1): $X_{n+m/2+1}, \dots, X_{n+m} \sim \mathcal{N}(-\mu + \Delta, I)$

Consider the class 0 which is comprised of $n/2$ in-distribution and $m/2$ OOD samples.

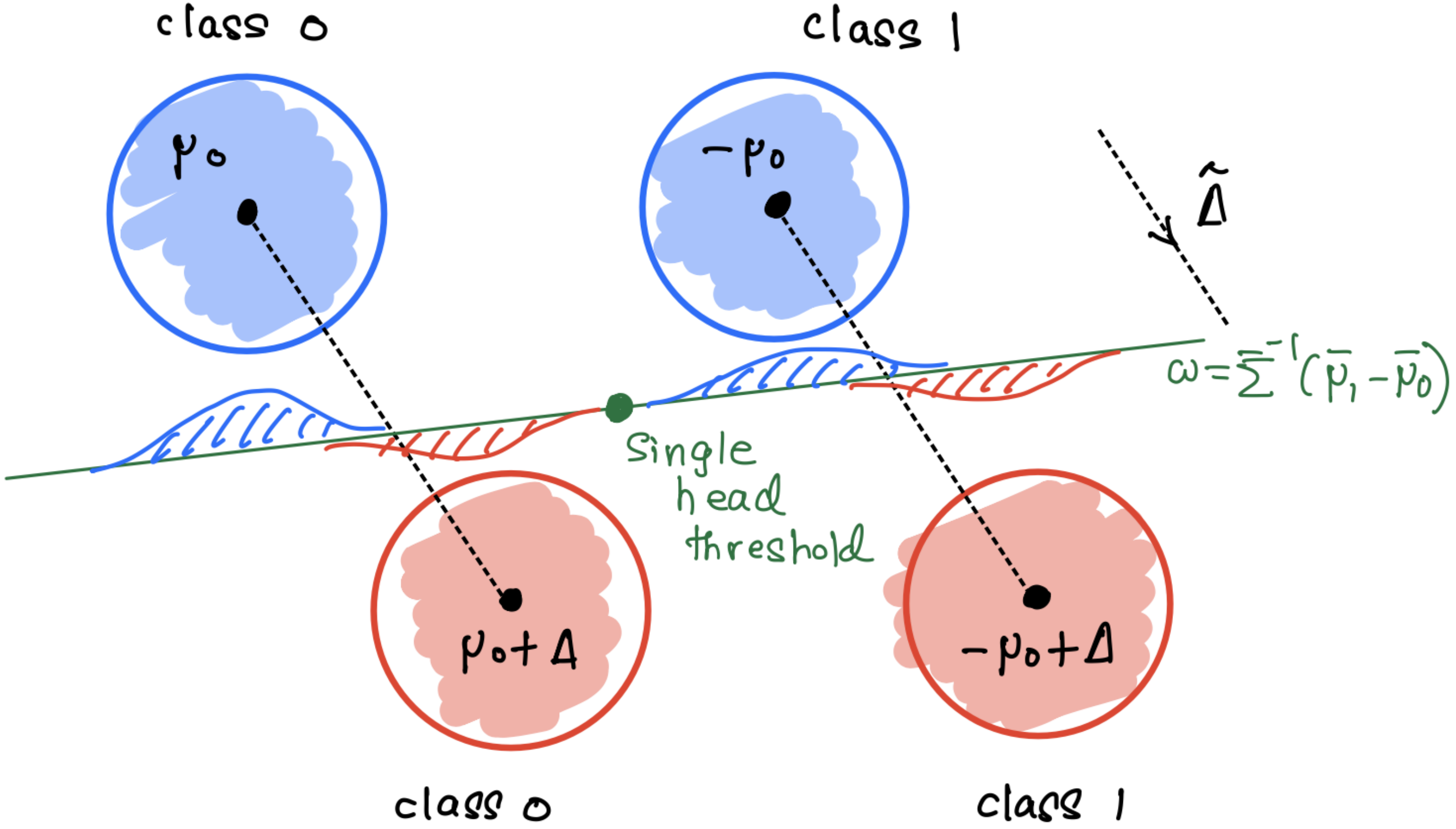
Let $\bar{\mu}_0$ and $\bar{\Sigma}_0$ be sample mean and sample covariance matrix of class 0.

$$\bar{\mu}_0 \sim \mathcal{N}\big(\mu + \frac{m}{n+m}\Delta, \frac{1}{n+m}I\big)$$

Similarly,

$$\bar{\mu}_1 \sim \mathcal{N}\big(-\mu + \frac{m}{n+m}\Delta, \frac{1}{n+m}I\big)$$

# Bivariate Single-Head LDA



# Bivariate Single-Head LDA

The projection vector of the LDA is given by, (Assuming $\bar{\Sigma}_0 = \bar{\Sigma}_1 = \bar{\Sigma}$),

$$w = \bar{\Sigma}^{-1}(\bar{\mu}_1 - \bar{\mu}_0)$$

The threshold is given by,

$$c = w \cdot \frac{(\bar{\mu}_0 + \bar{\mu}_1)}{2}$$

Then the LDA classification rule is given by,

$$g(x) = \mathbb{I}(w \cdot x > c)$$

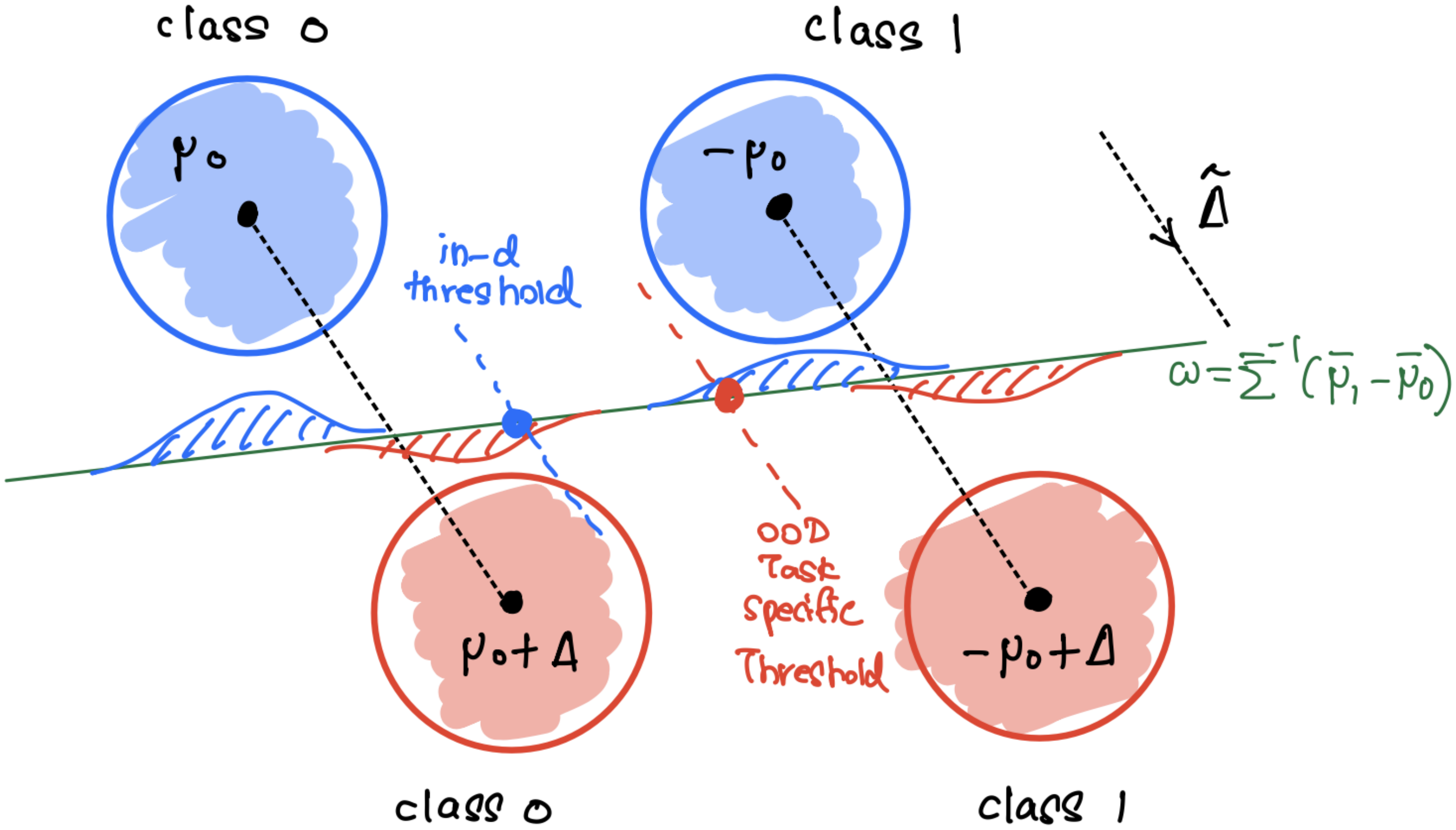
# Bivariate Single-Head LDA

The risk $L_{n,m,\Delta}$ on the in-distribution data is given by,

$$L_{n,m,\Delta} = \mathbb{E}_{f_{w,c}}[L(w,c)]$$

$$L(w,c) = \mathbb{P}_{x \sim f_0}[w \cdot x > c] + \mathbb{P}_{x \sim f_1}[w \cdot x < c]$$

# Bivariate Multi-Head LDA



# Bivariate Multi-Head LDA

The shared projection vector of the LDA is given by, (Assuming $\bar{\Sigma}_0 = \bar{\Sigma}_1 = \bar{\Sigma}$),

$$w = \bar{\Sigma}^{-1}(\bar{\mu}_1 - \bar{\mu}_0)$$

The task-specific thresholds are given by,

$$c_{in} = w \cdot \frac{(\bar{\mu}_{0,in} + \bar{\mu}_{1,in})}{2}; \quad \bar{\mu}_{0,in} \sim \mathcal{N}(\mu, I), \bar{\mu}_{1,in} \sim \mathcal{N}(-\mu, I)$$

$$c_{out} = w \cdot \frac{(\bar{\mu}_{0,out} + \bar{\mu}_{1,out})}{2}; \quad \bar{\mu}_{0,out} \sim \mathcal{N}(\mu + \Delta, I), \bar{\mu}_{1,out} \sim \mathcal{N}(-\mu + \Delta, I)$$

Then the LDA classification rule for in-distribution data is given by,

$$g(x) = \mathbb{I}(w \cdot x > c_{in})$$

## Bivariate Single-Head LDA

The risk $L_{n,m,\Delta}$ on the in-distribution data is given by,

$$L_{n,m,\Delta} = \mathbb{E}_{f_{w,c_{in}}}\left[L(w, c_{in})\right]$$

$$L(w, c_{in}) = \mathbb{P}_{x \sim f_0}[w \cdot x > c_{in}] + \mathbb{P}_{x \sim f_1}[w \cdot x < c_{in}]$$