Machine Learning 1 - Homework 2

Selene Baez Santamaria

1 MAP solution for Linear Regression

In class we solved for the maximum likelihood estimator for linear regression with polynomial basis functions. In this exercise you will solve for the *maximum a posterior* (MAP) solution: $\mathbf{w}_{\text{MAP}} = \left(\mathbf{\Phi}^T\mathbf{\Phi} + \lambda\mathbf{I}\right)^{-1}\mathbf{\Phi}^T\mathbf{t}$. For this problem we assume N training vectors $\{\mathbf{x}_n\}_{n=1}^N$, each of which is mapped using basis functions to $\boldsymbol{\phi}_n$. In the training set, the data come in input-output pairs, i.e. $\{\mathbf{x}_n, t_n\}$. We assume that one of the basis functions is the constant 1, and there are $M-1$ other basis function in $\boldsymbol{\phi}_n$. We also have the following information:

- The regression prediction: $y(\mathbf{x}_n, \mathbf{w}) = \mathbf{w}^T\boldsymbol{\phi}_n$.
- The likelihood function: $p(t_n|\boldsymbol{\phi}_n, \mathbf{w}, \beta) = \mathcal{N}\left(t_n|\mathbf{w}^T\boldsymbol{\phi}_n, 1/\beta\right)$
- The prior over $\mathbf{w}$: $p(\mathbf{w}) = \mathcal{N}\left(\mathbf{w}|\mathbf{0}, \mathbf{I}/\alpha\right)$. $\mathbf{I}$ is the identity matrix, $\mathbf{0}$ is a vector of 0's.
- The data are iid (independently and identically distributed).

Answer the following:

1. Write down the likelihood $p(\mathcal{D}|\boldsymbol{\theta})$ using a) a product over N and b) in vector/matrix form. Tip: You can answer both a) and b) in one set of equations by starting with a), then simplifying to get b). For b) make sure to define any matrices and vectors.
   Solution:

$$
\begin{aligned}
p(\mathcal{D}|\boldsymbol{\theta}) &= \prod_{n=1}^{N} \mathcal{N}(t_n|\mathbf{w}^T\boldsymbol{\phi}_n, 1/\beta) \\
&= \prod_{n=1}^{N} \frac{\beta^{1/2}}{(2\pi)^{1/2}} \exp(-\frac{\beta}{2}(t_n - \mathbf{w}^T\boldsymbol{\phi}_n)^2) \\
&= \frac{\beta^{N/2}}{(2\pi)^{N/2}} \prod_{n=1}^{N} \exp(-\frac{\beta}{2}(t_n - \mathbf{w}^T\boldsymbol{\phi}_n)^2) \\
&= \frac{\beta^{N/2}}{(2\pi)^{N/2}} \exp(-\frac{\beta}{2}\sum(t_n - \mathbf{w}^T\boldsymbol{\phi}_n)^2) \\
&= \frac{\beta^{N/2}}{(2\pi)^{N/2}} \exp(-\frac{\beta}{2}(\mathbf{t} - \mathbf{\Phi}\mathbf{w})^T(\mathbf{t} - \mathbf{\Phi}\mathbf{w})) \\
&= \mathcal{N}(\mathbf{t}|\mathbf{\Phi}\mathbf{w}, \beta^{-1}\mathbf{I})
\end{aligned}
$$

2. Write down the prior $p(\mathbf{w})$ (by expanding the expression for multivariate Gaussian distribution). Compute its log.
Solution:

$$
\begin{aligned}
p(\mathbf{w}) &= \mathcal{N}(\mathbf{w}|\mathbf{0}, \frac{1}{\alpha}\mathbf{I}) \\
&= \frac{\alpha^{D/2}}{(2\pi)^{D/2}} \exp(-\frac{\alpha}{2}\mathbf{w}^T\mathbf{w})
\end{aligned}
$$

$$
\begin{aligned}
\ln p(\mathbf{w}) &= \frac{D}{2}\ln\alpha - \frac{D}{2}\ln(2\pi) - \frac{\alpha}{2}\mathbf{w}^T\mathbf{w} \\
&= \frac{\alpha}{2}\mathbf{w}^T\mathbf{w} + \mathcal{C}
\end{aligned}
$$

3. Write down an expression for the posterior over $\mathbf{w}$. Remember this will involve applying Bayes rule to the prior, likelihood, and evidence. The evidence will require an integral. You do not need the analytic form for the evidence, but you need the correct variables and conditioning variables, e.g. something like $p(a|b, c)$ where you define a, b, and c.
Solution:

$$
\begin{aligned}
p(\mathbf{w}|\mathcal{D}) &= \frac{\mathcal{N}(\mathbf{w}|\mathbf{0}, \frac{1}{\alpha}\mathbf{I}) \prod_{n=1}^{N} \mathcal{N}(t_n|\mathbf{w}^T\boldsymbol{\phi}_n, 1/\beta)}{\int \mathcal{N}(\mathbf{w}|\mathbf{0}, \frac{1}{\alpha}\mathbf{I}) \prod_{n=1}^{N} \mathcal{N}(t_n|\mathbf{w}^T\boldsymbol{\phi}_n, 1/\beta) d\mathbf{w}} \\
&= \frac{\mathcal{N}(\mathbf{w}|\mathbf{0}, \frac{1}{\alpha}\mathbf{I})\mathcal{N}(\mathbf{t}|\boldsymbol{\Phi}\mathbf{w}, \frac{1}{\beta}\mathbf{I})}{\int \mathcal{N}(\mathbf{w}|\mathbf{0}, \frac{1}{\alpha}\mathbf{I})\mathcal{N}(\mathbf{t}|\boldsymbol{\Phi}\mathbf{w}, \frac{1}{\beta}\mathbf{I}) d\mathbf{w}} \\
&= \frac{\mathcal{N}(\mathbf{w}|\mathbf{0}, \frac{1}{\alpha}\mathbf{I})\mathcal{N}(\mathbf{t}|\boldsymbol{\Phi}\mathbf{w}, \frac{1}{\beta}\mathbf{I})}{p(t|\boldsymbol{\Phi}, \alpha, \beta)}
\end{aligned}
$$

4. Compute the log-posterior, both for the a) and b) likelihood forms from above. Collect everything that does not depend on **w** into a constant C. What parts of the previous expression do not depend on **w**? Why is finding the MAP much simpler than finding the full posterior distribution?
Solution:

$$
\begin{aligned}
\ln p(\mathbf{w}|\mathcal{D}) &= -\frac{\alpha}{2}\mathbf{w}^T\mathbf{w} - \frac{\beta}{2}\sum_{n=1}^{N}(t_n - \mathbf{w}^T\boldsymbol{\phi}_n)^2 + \mathcal{C} \\
&= \frac{\alpha}{2}\mathbf{w}^T\mathbf{w} - \frac{\beta}{2}(\mathbf{t} - \boldsymbol{\Phi}\mathbf{w})^T(\mathbf{t} - \boldsymbol{\Phi}\mathbf{w}) + \mathcal{C}
\end{aligned}
$$

5. Solve for $\mathbf{w}_{\text{MAP}}$ by a) taking the derivative of the log-posterior with respect to **w**, b) setting it to 0, and c) solving for **w**. Do this for both forms of likelihood.
Solution:

$$
\begin{aligned}
\ln p(\mathbf{w}|\mathcal{D}) &= \frac{\alpha}{2}\mathbf{w}^T\mathbf{w} - \frac{\beta}{2}\sum_{n=1}^{N}(t_n - \mathbf{w}^T\boldsymbol{\phi}_n)^2 + \mathcal{C} \\
\frac{\partial \ln p(\mathbf{w}|\mathcal{D})}{\partial \mathbf{w}} &= -\alpha\mathbf{w} - \beta\sum_{n=1}^{N}(t_n - \mathbf{w}^T\boldsymbol{\phi}_n)(-\boldsymbol{\phi}_n) = 0 \\
\alpha\mathbf{w} &= \beta\sum_{n=1}^{N}(t_n - \mathbf{w}^T\boldsymbol{\phi}_n)\boldsymbol{\phi}_n \\
\alpha\mathbf{w} &= \beta\sum_{n=1}^{N} t_n\boldsymbol{\phi}_n - \underbrace{\mathbf{w}^T\boldsymbol{\phi}_n}_{\text{scalar}}\boldsymbol{\phi}_n \\
\alpha\mathbf{w} &= \beta\sum_{n=1}^{N} t_n\boldsymbol{\phi}_n - \boldsymbol{\phi}_n \quad \underbrace{\mathbf{w}^T\boldsymbol{\phi}_n}_{\text{same as } \boldsymbol{\phi}_n^T\mathbf{w}} \\
(\alpha\mathbf{I} + \beta\sum_{n=1}^{N}\boldsymbol{\phi}_n\boldsymbol{\phi}_n^T)\mathbf{w} &= \beta\sum_{n=1}^{N} t_n\boldsymbol{\phi}_n
\end{aligned}
$$

$$
\mathbf{w}_{\text{MAP}} = (\alpha\mathbf{I} + \beta\sum_{n=1}^{N}\boldsymbol{\phi}_n\boldsymbol{\phi}_n^T)^{-1}\beta\sum_{n=1}^{N} t_n\boldsymbol{\phi}_n
$$

$$
\begin{aligned}
\ln p(\mathbf{w}|\mathcal{D}) &= -\frac{\alpha}{2}\mathbf{w}^T\mathbf{w} - \frac{\beta}{2}(\mathbf{t} - \boldsymbol{\Phi}\mathbf{w})^T(\mathbf{t} - \boldsymbol{\Phi}\mathbf{w}) + \mathcal{C} \\
&= -\frac{\alpha}{2}\mathbf{w}^T\mathbf{w} - \frac{\beta}{2}\mathbf{w}^T\boldsymbol{\Phi}^T\boldsymbol{\Phi}\mathbf{w} + \beta\mathbf{w}^T\boldsymbol{\Phi}^T\mathbf{t} + \mathcal{D}
\end{aligned}
$$

$$
\frac{\partial \ln p(\mathbf{w}|\mathcal{D})}{\partial \mathbf{w}} = -\alpha\mathbf{w} - \beta\boldsymbol{\Phi}^T\boldsymbol{\Phi}\mathbf{w} + \beta\boldsymbol{\Phi}^T\mathbf{t} = 0
$$

$$
(\alpha\mathbf{I} + \beta\boldsymbol{\Phi}^T\boldsymbol{\Phi})\mathbf{w} = \beta\boldsymbol{\Phi}^T\mathbf{t}
$$

$$
\begin{aligned}
\mathbf{w}_{\text{MAP}} &= (\alpha\mathbf{I} + \beta\boldsymbol{\Phi}^T\boldsymbol{\Phi})^{-1}\beta\boldsymbol{\Phi}^T\mathbf{t} \\
&= (\beta(\frac{\alpha}{\beta}\mathbf{I} + \boldsymbol{\Phi}^T\boldsymbol{\Phi}))^{-1}\beta\boldsymbol{\Phi}^T\mathbf{t} \\
&= \frac{1}{\beta}(\lambda\mathbf{I} + \boldsymbol{\Phi}^T\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}^T\mathbf{t}
\end{aligned}
$$

6. **BONUS** Our prior for $\mathbf{w}$ assumes the same marginal distribution for each entry in $\mathbf{w}$, including that of the first basis function $\boldsymbol{\phi}_0 = 1$. What is the role this basis function? Why should we avoid placing the same penalty/prior for this basis? Rewrite $p(\mathbf{w})$ so that the first basis function has its own prior/penalty.
   Solution:
   The constant basis function acts as a bias or offset for the regression problem. If we use the same prior for this weight as for the others, we are assuming that the offset from the y-axis should somehow be penalized. This does not make too much sense a priori, so instead we use a different precision for this basis function, i.e. $\alpha_0 << \alpha$, while using α for all the others.

2 Probability distributions, likelihoods, and estimators

For these questions you will be working with different probability density functions listed in the table below. The purpose of these questions is to practice working with a variety of PDFs and to make computing likelihoods, MLEs, etc. more natural. Note below the *indicator* notation $[x = 0]$ (and $[x = 1]$). The square brackets evaluate to 1 if the argument is true, and 0 otherwise. E.g. if x is 1, the $[x = 0] = 0$ and $[x = 1] = 1$ (here $[x = 0]$ is lazy notation; in Python you would write $x == 0$, for example). We will use the notation a lot, both below and when we learn about classification.

Distribution	$p(x\|\theta)$	Range of x	Range of θ
Bernouilli	$\theta^{[x=1]}(1-\theta)^{[x=0]}$	$x \in \{0, 1\}$	$0 \leq \theta \leq 1$
Beta	$\frac{\Gamma(\theta_1+\theta_0)}{\Gamma(\theta_1)\Gamma(\theta_0)} x^{\theta_1-1}(1-x)^{\theta_0-1}$	$0 \leq x \leq 1$	$\theta_1 > 0, \theta_0 > 0$
Poisson	$\frac{\theta^x}{x!} e^{-\theta}$	$x \in \{0, 1, 2, \ldots\}$	$\theta > 0$
Gamma	$\frac{\theta_1^{\theta_0}}{\Gamma(\theta_0)} x^{\theta_0-1} e^{-\theta_1 x}$	$x \geq 0$	$\theta_1 \geq 0$, $\theta_0 \geq 0$
Gaussian	$\frac{1}{\sqrt{2\pi\theta_1}} e^{-\frac{1}{2}\left(\frac{x-\theta_0}{\theta_1}\right)^2}$	$-\infty < x < \infty$	$-\infty < \theta_0 < \infty$, $\theta_1 > 1$

Question 2.1

For each of the probability distributions above, write down their normalizing constants. Remember that $\int p(x|\theta)dx = 1$ for continuous x and $\sum_x p(x|\theta) = 1$ for discrete x.

Question 2.2

You live in Amsterdam and find that it rains quite a lot. You want to estimate the probability that it will rain any given day of the year. Every month for a year you count the number of days with rain, and you get the following (from January to December): 22,19,16,16,14,14,17,18,19,20,21,21 (for a grand total of 217 days with rain).[1] Let r_t be an observation for day t in the year; $r_t = 1$ means there was some rain on day t, $r_t = 0$ means there was no rain. We want to estimate the parameter ρ, the probability of rain on any day of the year. We assume a Bernouilli distribution for the observations $\{r_t\}_{t=1}^{365}$, that is $p(r_t|\rho) = \text{Bernouilli}(r_t|\rho)$. To answer these questions, the number of days of rain per month is not important, only the total for the year is relevant. With this information, answer the following questions:

1. What is the likelihood for a single observation? For the entire set of observations?
   Solution:

$$p(r_t|\rho) = \rho^{r_t}(1-\rho)^{1-r_t}$$

$$\begin{aligned} p(\mathbf{r}|\rho) &= \prod_{t=1}^{T} \rho^{r_t}(1-\rho)^{1-r_t} \\ &= \rho^{\sum_{t=1}^{T} r_t}(1-\rho)^{\sum_{t=1}^{T} 1-r_t} \\ &= \rho^{n_1}(1-\rho)^{n_0} \end{aligned}$$

2. Write the log-likelihood for the entire set of observations.
   Solution:

$$\ln p(\mathbf{r}|\rho) = n_1 \ln \rho + n_0 \ln(1-\rho)$$

3. Solve for the MLE of ρ. Do it in general (with symbols for counts n_0, n_1 for days without and with rain) and for this specific case (plug-in the numbers).
   Solution:

[1] Source: `http://www.amsterdam.climatemps.com/`.

$$\begin{aligned}\frac{\partial \ln p(\mathbf{r}|\rho)}{\partial \rho} &= \frac{n_1}{\rho} + \frac{n_0}{1-\rho}(-1) = 0\\ \frac{n_1}{\rho} &= \frac{n_0}{1-\rho}\\ n_1 - n_1\rho &= n_0\rho\\ \rho &= \frac{n_1}{N}\\ &= 217/365\end{aligned}$$

4. Assume a Beta prior for ρ with parameters a and b. What is the MAP for ρ?
   Solution:

$$\begin{aligned}f = \ln p(\rho|\mathbf{r}) &\propto \ln p(\rho|\mathbf{r}) + \ln p(\rho)\\ &= n_1 \ln \rho + n_0 \ln(1-\rho) + (a-1)\ln \rho + (b-1)\ln(1-\rho)\end{aligned}$$

$$\begin{aligned}\frac{\partial f}{\partial \rho} &= \frac{n_1}{\rho} - \frac{n_0}{1-\rho} + \frac{a-1}{\rho} - \frac{b-1}{1-\rho} = 0\\ \rho &= \frac{n_1 + a - 1}{N + a + b - 2}\end{aligned}$$

5. Write the form of the posterior distribution for ρ? You do not need to solve it analytically.
   Solution:

$$\begin{aligned}p(\rho|\mathbf{r}) &= p(\mathbf{r}|\rho)p(\rho)\\ &= \frac{\rho^{n_1+a-1}(1-\rho)^{n_0+b-1}}{\int \rho^{n_1+a-1}(1-\rho)^{n_0+b-1}dp}\\ &= \frac{\Gamma(\mathcal{N}+a+b)}{\Gamma(n_1+a)\Gamma(n_0+b)}\rho^{n_1+a-1}(1-\rho)^{n_0+b-1}\\ &= \mathcal{B}(\rho|a+n_1, b+n_0)\end{aligned}$$

6. (Optional) Solve for the posterior distribution analytically. Hint: it is a Beta distribution.
   Solution:

Question 2.3
You work in the staffing department of a maternity hospital and part of your job is to determine the staffing requirements during the night shift at your hospital. This might mean the number of doctors and nurses at the hospital and the number of doctors on call (if there are more than the average number of deliveries). Your goal is to determine the distribution over the number of deliveries during the night shift $d_t \in \{0, 1, 2, \ldots\}$ (d for delivery count, t for time, the index of the night). With this you can compute the mean, the probability of more than 5 deliveries, etc. You collect data for two weeks, i.e. $d_1, \ldots, d_{14} = 4, 7, 3, 0, 2, 2, 1, 5, 4, 4, 3, 3, 2, 3$. You assume the observations are explained by a Poisson distribution with parameter λ over the discrete delivery counts. With this information, answer the following questions:

1. What is the likelihood for a single observation? For the entire set of observations?
   Solution:

$$p(d_t|\lambda) = \frac{\lambda^{d_t}}{d_t!}\exp(-\lambda)$$

$$\begin{aligned} p(\mathbf{d}|\lambda) &= \prod_{t=1}^{T} \frac{\lambda^{d_t}}{d_t!}\exp(-\lambda) \\ &= \frac{\lambda^{\sum_{t=1}^{T} d_t}}{\prod_{t=1}^{T} d_t!}\exp(-T\lambda) \\ &= \frac{\lambda^{n}}{\prod_{t=1}^{T} d_t!}\exp(-T\lambda)) \end{aligned}$$

2. Write the log-likelihood for the entire set of observations.
   Solution:

3. Solve for the MLE of λ. Do it in general and for this specific case (plug-in the numbers).
   Solution:

4. Assume a Gamma prior for λ with parameters a and b. What is the MAP estimate of λ?
   Solution:

5. Write the form of the posterior distribution for λ? (You do not need to solve it analytically)
Solution:

6. (Optional) Solve for the posterior distribution analytically. Hint: it is a Gamma distribution.
Solution:

Question 2.4

You have developed a blood test aimed at detecting a disease $d \in \{0, 1\}$ (disease is absent ($d = 0$) or present ($d = 1$)). The test measures the level of a specific indicator of the disease, that is it returns a real valued number relative to some baseline (so the levels can be both negative and positive – anywhere along the real line). Two models of the population are built: one for the patients with the disease, and another for the general population. Measurements tend to have a Gaussian shape, and we therefore model the entire population as a mixture of two Gaussians. That is, $p(l) = p(d = 0)p(l|d = 0) + p(d = 1)p(l|d = 1)$, where $p(d)$ is the prior distribution of patients with and without the disease in the general population and $p(l|d)$ are conditional Gaussian distributions, one for the patients with disease, and one for those without. Note: with this question and the previous two, we are simply applying rules of probability (with some algebra) to get the form of the posterior distribution; however, in this problem we are also classifying (since our target is the discrete label d).

Assume we know $p(d = 0) = \pi_0 = 0.999$ and $p(d = 1) = \pi_1 = 0.001$ from previous experience. We do not know the parameters μ_0, σ_0^2 (the mean and variance of the disease-free population) nor μ_1, σ_1^2 (for the disease population). We measure levels $\{l_n\}_{n=1}^{N}$ for N people, and we know that $n \in \{D_0\}$ are the indices for the disease free patients and $n \in \{D_1\}$ are the indices for the patients with the disease (i.e. D_0 and D_1 are non-intersecting sets of indices from 1 to N). With this information, answer the following questions:

1. Write down the likelihood of the observations as a product over N level recordings. Hint: use indicator notation (like in the Bernouilli distribution) to distinguish between $d_n = 0$ and $d_n = 1$ in the likelihood.
Solution:

$$p(\mathbf{d}, \mathbf{l}|\boldsymbol{\theta}) = \prod_{n=1}^{N} (\pi_1 \mathcal{N}(l_n|\mu_1, \sigma_1^2))^{d_n} (\pi_0 \mathcal{N}(l_n|\mu_0, \sigma_0^2))^{1-d_n}$$

2. Write down the likelihood as a product over the likelihoods for $\{D_0\}$ and $\{D_1\}$.
Solution:

$$p(\mathbf{d}, \mathbf{l}|\boldsymbol{\theta}) = (\prod_{n \in D_1} \pi_1 \mathcal{N}(l_n|\mu_1, \sigma_1^2))(\prod_{n \in D_0} \pi_0 \mathcal{N}(l_n|\mu_0, \sigma_0^2))$$

3. Compute the log-likelihood.
Solution:

$$\begin{aligned}
\ln p(\mathbf{d}, \mathbf{l}|\boldsymbol{\theta}) &= \sum_{n \in D_1} (\ln \pi_1 + \ln \mathcal{N}(l_n|\mu_1, \sigma_1^2)) + \sum_{n \in D_0} (\ln \pi_0 + \ln \mathcal{N}(l_n|\mu_0, \sigma_0^2)) \\
&= \sum_{n \in D_1} (\ln \pi_1 + \frac{1}{2}\ln(2\pi) - \frac{1}{2}\ln(\sigma_1^2) - \frac{1}{2}\frac{(l_n - \mu_1)^2}{\sigma_1^2}) \\
&+ \sum_{n \in D_0} (\ln \pi_0 + \frac{1}{2}\ln(2\pi) - \frac{1}{2}\ln(\sigma_0^2) - \frac{1}{2}\frac{(l_n - \mu_1)^2}{\sigma_0^2})
\end{aligned}$$

4. Find the MLE for μ_0 and σ_0^2. Assume we can do the same for μ_1 and σ_1^2
Solution:

$$\begin{aligned}
\frac{\partial \ln p(\mathbf{d}, \mathbf{l}|\boldsymbol{\theta})}{\partial \mu_0} &= \sum_{n \in D_0} (-\frac{(l_n - \mu_0)}{\sigma_0^2}(-1)) = 0 \\
0 &= \sum_{n \in D_0} (l_n - \mu_0) \\
0 &= \sum_{n \in D_0} l_n - N_0 \mu_0 \\
\mu_0 &= \frac{1}{N_0} \sum_{n \in D_0} l_n
\end{aligned}$$

$$\begin{aligned}
\frac{\partial \ln p(\mathbf{d}, \mathbf{l}|\boldsymbol{\theta})}{\partial \sigma_0^2} &= \sum_{n \in D_0} (-\frac{1}{2\sigma_0^2} - \frac{1}{2}\frac{(l_n - \mu_0)^2}{(\sigma_0^2)^2}(-1)) \\
\frac{N_0}{2\sigma_0^2} &= \frac{1}{2}\frac{\sum_{n \in D_0}(l_n - \mu_0)^2}{(\sigma_0^2)^2} \\
\sigma_0^2 &= \frac{1}{N_0} \sum_{n \in D_0} (l_n - \mu_0)^2
\end{aligned}$$

5. We now have our models. To make a prediction, solve for $p(d = 1|l_\star)$, where $l_\star$ is a level recorded for a new patient. Hint: use Bayes theorem.
   Solution:

$$\begin{aligned} p(d = 1|l_*) &= \frac{\pi_1 \mathcal{N}(l_*|\mu_1, \sigma_1^2)}{\pi_1 \mathcal{N}(l_*|\mu_1, \sigma_1^2) + \pi_0 \mathcal{N}(l_*|\mu_0, \sigma_0^2)} \\ &= \frac{1}{1 + \frac{\pi_0 \mathcal{N}(l_*|\mu_0, \sigma_0^2)}{\pi_1 \mathcal{N}(l_*|\mu_1, \sigma_1^2)}} \\ &= \frac{1}{1 + \exp^{-a(l_*)}} \\ a(l_*) &= \ln(\frac{\pi_1 \mathcal{N}(l_*|\mu_1, \sigma_1^2)}{\pi_0 \mathcal{N}(l_*|\mu_0, \sigma_0^2)}) \end{aligned}$$

6. Reduce your solution to have the form of a sigmoid, i.e.

$$p(d = 1|l_\star) = \frac{1}{1 + e^{-a(l_\star)}}.$$

Solution:

$$p(d = 1|l_*) = \frac{1}{1 + \exp^{-a(l_*)}} = \frac{1}{1 + \frac{\pi_0 \mathcal{N}(l_*|\mu_0, \sigma_0^2)}{\pi_1 \mathcal{N}(l_*|\mu_1, \sigma_1^2)}}$$