1.1. Consider a univariate continuous measurement x and use it to predict a univariate continuous state w (i.e. regression). Think of two computer vision tasks that accomplish this.

(a) Determining the height/width/dimensions of something based off the a picture with a fixed reference point.

(b) Determining the best path to get someone out of rubble (during natural disaster scenarios) by mapping out the environment.

2.1. Consider the case where the observed measurement x is univariate and continuous, but the world state w is discrete and can take one of two values ($w \in \{0, 1\}$). Think of two computer vision tasks that accomplish this.

(a) Predict whether or not an medical image/scan contains a tumor/disease.

(b) Classifying a parts of the image as background or foreground.

1.2. Suppose we use Gaussian distributions for both likelihood and prior

(a) Write down the concrete formula in terms of PDF functions for them.

i. Likelihood: $p(x \mid w) = N_{x_i}(\phi_0 + \phi_1 w, \sigma^2)$

ii. Prior: $p(w) = N_{w_i}(\mu_p, \sigma_p^2)$

(b) For the unknown parameters in the likelihood and prior (i.e. mean and variance), compute the MLE solutions for them respectively in detailed steps with clear definitions of notations.

i. Likelihood:
Easier to work with the log of this function, since the logarithm is a monotonic function, the maximum will still be the same. So,

$$\hat{\phi_0} + \hat{\phi_1}\hat{w}, \hat{\sigma}^2 \quad = \quad \underset{\phi_0+\phi_1 w, \sigma^2}{\operatorname{argmax}} \left[\log\left[N_{x_i}(\phi_0 + \phi_1 w, \sigma^2)\right]\right]$$

where N_{x_i} is the Gaussian distribution for each data point $x_i \dots x_n$.

$$= \quad \underset{\phi_0+\phi_1 w, \sigma^2}{\operatorname{argmax}} \left[\log\left[\frac{1}{\sqrt{2\pi\sigma^2}}\right] - \sum_{i=1}^{n} \frac{(x_i - (\phi_0 + \phi_1 w))^2}{2\sigma^2}\right]$$

which gets reduced to

$$= \quad \underset{\phi_0+\phi_1 w, \sigma^2}{\operatorname{argmax}} \left[-0.5\log[2\pi] - 0.5\log\sigma^2 - \sum_{i=1}^{n} \frac{(x_i - (\phi_0 + \phi_1 w))^2}{2\sigma^2}\right]$$

Maximize according to mean first, meaning set derivative w.r.t. mean to 0

$$\frac{\partial}{\partial(\phi_0 + \phi_1 w)} = \sum_{i=1}^{n} \frac{(x_i - (\phi_0 + \phi_1 w))}{\sigma^2} = 0$$

to find that

$$\hat{\phi_0} + \hat{\phi_1}\hat{w} = \frac{\sum_{i=1}^{n} x_i}{n}$$

and using a similar process with the variance, we find

$$\frac{\partial}{\partial \sigma} = \sum_{i=1}^{n} \frac{(x_i - (\phi_0 + \phi_1 w))^2}{n\sigma^2} - 1 = 0$$

to find that

$$\hat{\sigma}^2 = \sum_{i=1}^{n} \frac{(x_i - (\hat{\phi_0} + \hat{\phi_1}\hat{w}))^2}{n}$$

ii. Prior: Similar to previous problem

$$\hat{\mu_p}, \hat{\sigma_p}^2 = \underset{\mu_p, \sigma_p^2}{\operatorname{argmax}} \left[\log\left[N_{w_i}(\mu_p, \sigma_p^2)\right]\right]$$

$$= \underset{\mu_p, \sigma_p^2}{\operatorname{argmax}} \left[-0.5\log[2\pi] - 0.5\log\sigma_p^2 - \sum_{i=1}^{n} \frac{(w_i - \mu_p)^2}{2\sigma_p^2}\right]$$

Maximize w.r.t. the mean to get

$$\sum_{i=1}^{n} \frac{(w_i - \mu_p)}{\sigma_p^2} = 0$$

to find that

$$\hat{\mu_p} = \frac{\sum_{i=1}^{n} w_i}{n}$$

Maximize w.r.t. standard deviation to also find

$$\hat{\sigma_p}^2 = \sum_{i=1}^{n} \frac{(w_i - \hat{\mu_p})^2}{n}$$

(c) Consider the prior $p(w)$, assume only the mean parameter is unknown. Compute the MAP solution for it in detailed steps with clear definitions of notations, assuming the prior distribution of the unknown mean (as r.v.) is a Gaussian distribution with known mean and variance.

Calculating the MAP is similar to MLE, but we use

$$\underset{\theta}{\operatorname{argmax}} \left[P_{\theta|x}(\theta \mid x)\right] = \underset{\theta}{\operatorname{argmax}} \left[P_{x|\theta}(x \mid \theta) P_\theta(\theta)\right]$$

and since we already know

$$P_{x|\theta}(x \mid \theta) = p(x \mid w) = \text{MLE}$$

then applying a natural logarithm, a monotonic function, we get

$$\underset{w}{\operatorname{argmax}} \left[p(x \mid w)p(w)\right] = \underset{w}{\operatorname{argmax}} \left[\log p(x \mid w) + \log p(w)\right]$$

which becomes

$$\frac{\partial}{\partial w}(\log p(x \mid w) + \log p(w)) = 0$$

$$\frac{\partial}{\partial w}\left(\log\left(N_{x_i}(\phi_0 + \phi_1 w, \sigma^2)\right) + \log\left(N_{w_i}(\mu_p, \sigma_p^2)\right)\right)$$

and can further be simplified.

2.2. Suppose we use Bernoulli distribution for the posterior.

(a) Write down the concrete formula for $p(w \mid x)$ in terms of the PMF function.
Posterior:

$$p(w \mid x) = \text{Bern}_w[sig[\phi_0 + \phi_1 x]]$$

where sig is the sigmoid function and $w \in \{0, 1\}$

(b) For the unknown parameters in the posterior you defined, compute the MLE solutions in detailed steps with clear definitions of notations. You may not have closed form solutions, which is fine as long as you show detailed steps.

$$\underset{\phi_0+\phi_1 x}{\text{argmax}} \left[\text{Bern}_w \left[\frac{1}{1 + \exp\left[-(\phi_0 + \phi_1 x)\right]} \right] \right]$$

$$\underset{\phi_0+\phi_1 x}{\text{argmax}} \left[\sum_{i=1}^{n} \left[\frac{1}{1 + \exp\left[-(\phi_0 + \phi_1 x)\right]} \right]^{w_i} \left(\left[\frac{1}{1 + \exp\left[-(\phi_0 + \phi_1 x)\right]} \right] \right)^{(1-w_i)} \right]$$

Similar to previous problem, take the natural logarithm and set derivative w.r.t. the Bernoulli input to 0, to get

$$\frac{1}{1 + \exp\left[-(\hat{\phi_0} + \hat{\phi_1}\hat{x})\right]} = \frac{1}{n} \sum_{i=1}^{n} w_i$$

which further gets simplified to

$$\hat{\phi_0} + \hat{\phi_1}\hat{x} = -\log\left(\frac{n}{\sum_{i=1}^{n} w_i} - 1 \right)$$

3. Consider the cat binary classification example again, propose your ideas of how to handle the challenges (viewpoint, illumination, etc) in the framework you choose (generative or discriminative or both), try to frame your proposal in terms of data collection, representation schema, modeling paradigm, learning setup.

If we use a discriminative model with parameters which include taking into account lighting, angle, etc., images with similar characteristics can be compared to determine whether or not a cat is in the image. The model would be flexible regarding the values of these parameters, and grouping images based on similar parameters would get similar looking images. In non-classical machine learning, like deep learning, a human face/head can have a generative model with parameters such as length of hair, colour of hair, shape of face, masculine/feminine, etc., and this same approach can be utilized for the cat images.