ECBME 4040 Homework 1

Jingyu Qian (UNI jq2250)

In collaboration with Wenyu Fu (UNI wf2223)

Problem c

Algorithm: minibatch SGD

Data: Training set X

for *i to 1 to N* **do**

    /* Select a random batch of training examples */

    $X_m \leftarrow$ *a random example sets of X with size* $(s_m, 65536)$

    /* Deduct previous $d_i s'$ contributions for this sample part */

    $A_i \leftarrow X_m^T X_m - \sum_{j=0}^{i-1} \lambda_j d_j d_j^T$

    initialize d_i *and let* $t = 1$

    /* ϵ is the stopping criterion in the following */

    while $t \leq T$ *and* $||\Delta d_i|| > \epsilon$ **do**

        $d_i \leftarrow d_i - \eta \cdot \nabla_{d_i}(d_i^T A_i d_i)$

        normalize d_i

        $t \leftarrow t + 1$

    end

    $\lambda_i \leftarrow d_i^T X^T X d_i$

end

Problem d

Answer

1.

In histogram equalization, after the transformation $y = g(x)$ y remains to be a constant, in this case, $\frac{1}{255}$ for all possible gray-scale values. The number of pixels mapped from x to y should be unchanged, *i.e.*

$$P(x)dx = P(y)dy$$

Consider in the $0 - 255$ range, $p(y) = \frac{1}{255}$, so we have:

$$\frac{dy}{dx} = 255 \cdot p(x)$$

The left side is the derivative of the function $y = g(x)$ that we seek. so integrating the right side:

$$y = g(x) = 255 \int_0^x p(x)dx \approx 255 \int_0^x f(x)dx = 255 \cdot cdf(f(x))$$

Here *cdf* means cumulative distribution function, which is the integral of the probability density function $f(x)$. So we have:

$$g(x) = 255 \cdot \frac{1}{\sigma\sqrt{2\pi}} \int_0^x exp(-\frac{(t-\mu)^2}{2\sigma^2})dt$$

2.

The expression given is symmetric over x, y, z.
Solving $P(X = x)$:

$$P(X = x) = \int_0^1 \int_0^1 8xyzdydz = 2x \quad x \in [0, 1]$$

For x outside $[0, 1]$, $p = 0$. The same could be applied to $p(Y = y)$ and $p(Z = z)$:

$$P(Y = y) = \begin{cases} 2y & y \in [0,1] \\ 0 & else \end{cases}, P(z = z) = \begin{cases} 2z & z \in [0,1] \\ 0 & else \end{cases}$$

For the expectation we have:

$$E(XYZ) = E(X)E(Y)E(Z) = 1 \cdot 1 \cdot 1 = 1$$

For the conditionally independent part:

$$P(X \cap Y|Z = z_0) = P\{(X, Y)|Z = z_0\} = \frac{P(X, Y, z_0)}{P(Z = z_0)} = \frac{8xyz_0}{z_0} = 4xy$$

$$P(X|Z = z_0) = \frac{P(X, z_0)}{P(Z = z_0)} = 2x$$

$$P(Y|Z = z_0) = \frac{P(Y, z_0)}{P(Z = z_0)} = 2y$$

$$\therefore P(X \cap Y|Z = z_0) = P(X|Z = z_0)P(Y|Z = z_0)$$

So X and Y are conditionally independent given Z. *i.e.* knowing Z makes X and Y independent.

Problem e

Answer

1.

$$f(x|\mu) = \frac{1}{(2\pi)^{n/2}} \cdot \frac{1}{|\Sigma|^{1/2}} exp[-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)]$$

$$P(\mu) = \frac{1}{(2\pi)^{n/2}} \cdot \frac{1}{|\Sigma_0|^{1/2}} exp[-\frac{1}{2}(\mu-\mu_0)^T\Sigma_0^{-1}(\mu-\mu_0)]$$

To get the maximum a posteriori we calculate:

$$\mu_{MAP}^T = \arg\max_\mu \frac{1}{(2\pi)^{nm/2}} \cdot \frac{1}{|\Sigma|^{m/2}} (\prod_{i=1}^m exp[-\frac{1}{2}(x^{(i)}-\mu)^T\Sigma^{-1}(x^{(i)}-\mu)]) \cdot \frac{1}{(2\pi)^{n/2}} \cdot \frac{1}{|\Sigma_0|^{1/2}} exp[-\frac{1}{2}(\mu-\mu_0)^T\Sigma_0^{-1}(\mu-\mu_0)]$$

$$= \arg\max_\mu (\prod_{i=1}^m exp[-\frac{1}{2}(x^{(i)}-\mu)^T\Sigma^{-1}(x^{(i)}-\mu)]) \cdot \exp[-\frac{1}{2}(\mu-\mu_0)^T\Sigma_0^{-1}(\mu-\mu_0)]$$

Focusing on the parameter of the exponential, and Calculating it's derivative:

$$-(\sum_{i=1}^m x^{(i)} - \mu)(2 \cdot \Sigma^{-1}) + (\mu - \mu_0)(2 \cdot \Sigma_0^{-1}) = 0$$

The maximum a posterior of μ is:

$$\mu_{MAP}^T = \frac{\sum_{i=1}^m x^{(i)T}\Sigma^{-1} + \mu_0^T\Sigma_0^{-1}}{\Sigma_0^{-1} + m\Sigma^{-1}}$$

Since the priori of Σ is not given, calculating the MAP is the same as calculating the MLE. With respect to Σ, the likelihood function is:

$$L(\Sigma) = \frac{1}{(2\pi)^{np/2}} \prod_{i=1}^m \frac{1}{|\Sigma|^{1/2}} exp[-\frac{1}{2}(x^{(i)}-\mu)^T\Sigma^{-1}(x^{(i)}-\mu)]$$

Similar to univariate situation, solving the likelihood functions we can get:

$$\Sigma_{MLE} = \sum_{i=1}^m (x^{(i)} - \bar{x})(x^{(i)} - \bar{x})^T$$

2.

$$E(\mu_{MAP}^T) = \frac{E(\sum_{i=1}^m x^{(i)T})\Sigma^{-1} + \mu_0^T\Sigma_0^{-1}}{\Sigma_0^{-1} + m\Sigma^{-1}}$$

$$= \frac{m\mu^T\Sigma^{-1} + \mu_0^T\Sigma_0^{-1}}{\Sigma_0^{-1} + m\Sigma^{-1}}$$

$$\because \lim_{m\to\infty} \frac{m\mu^T\Sigma^{-1} + \mu_0^T\Sigma_0^{-1}}{\Sigma_0^{-1} + m\Sigma^{-1}} = \mu^T$$

$\therefore \mu_{MAP}^T$ is asymptotically unbiased estimate of μ.

$$\because E(\Sigma_{MLE}) = \frac{1}{m}\Sigma$$

$$\lim_{m\to 0} \Sigma_{MLE} = \Sigma$$

$\therefore \Sigma_{MLE}$ is an asymptotically unbiased estimate of Σ.

3.

We can write Σ_{MLE} and Σ_{MAP} as:

$$E(\Sigma_{MLE}) = \sum_{i=1}^{m}(x^{(i)} - \mu_{MLE})(x^{(i)} - \mu_{MLE})^T$$

$$E(\Sigma_{MAP}) = \sum_{i=1}^{m}(x^{(i)} - \mu_{MAP})(x^{(i)} - \mu_{MAP})^T$$

since μ_{MAP} and MLE converge to the same number when number of observations m is large, we can say the maximum a posteriori estimate is asymptotically equal to maximum likelihood estimate.