First practicals in Machine learning 1 – 2024 – Paper 1

1 Calculus (September)

Find the derivatives of the following functions with respect to x.

(a) $\sigma(x) = \frac{1}{1+e^{-x}}$ (the standard logistic function or "sigmoid function").

(b) $\max\{0, x\}$ ("Rectified Linear Unit" or ReLu that is important in Deep Learning).

(c) What is the shape of the following derivative: $\frac{df(x)}{dx}$ $f : \mathbb{R} \to \mathbb{R}$, $x \in \mathbb{R}$

(d) What is the shape of the following derivative: $\frac{df(\boldsymbol{x})}{d\boldsymbol{x}}$ with $f : \mathbb{R}^n \to \mathbb{R}$, $\boldsymbol{x} \in \mathbb{R}^n$

(e) What is the shape of the following derivative: $\frac{d\boldsymbol{f}(\boldsymbol{x})}{d\boldsymbol{x}}$ with $\boldsymbol{f} : \mathbb{R}^n \to \mathbb{R}^m$, $\boldsymbol{x} \in \mathbb{R}^n$

(f) Calculate the following derivative $\frac{df(\mathbf{x})}{d\mathbf{x}}$ with $f(\mathbf{x}) = 2\exp(x_2 - \ln(x_1^{-1}) - \sin(x_3 x_1^2)), \mathbf{x} \in \mathbb{R}^3$.

(g) $\nabla_y h$ with $h = g(f(y))$ where $g(\mathbf{x}) = x_1^3 + \exp(x_2)$ and $\mathbf{x} := \mathbf{f}(y) = [y\sin(y), y\cos(y)]^T$. First show your understanding of the chain rule before plugging in the actual derivatives.

(h) We now assume that $\mathbf{x} := \mathbf{f}(y, z) = [y\sin(y) + z, y\cos(y) + z^2]^T$. Provide $\nabla_{y,z} h$. Hint: To determine the correct shape of $\nabla_{y,z} h$, view the input pair y and z as a vector $[y, z]^T$.

First practicals in Machine learning 1 – 2024 – Paper 1

2 Multivariate calculus (September)

The following questions are good practice in manipulating vectors and matrices and they are essential for solving for posterior distributions. Given the following expression:

$$(\mathbf{x} - \boldsymbol{\mu})^T \mathbf{S}^{-1} (\mathbf{x} - \boldsymbol{\mu})$$

where $\mathbf{x}$, $\boldsymbol{\mu}$, $\boldsymbol{\mu}_0$ are vectors and $\boldsymbol{\Sigma}^{-1}$ and $\mathbf{S}^{-1}$ are “symmetric”, “positive semi-definite” “invertible” matrices.

Answer the following questions:

(a) Expand the expression and gather terms.

(b) Collect all the terms that depend on $\boldsymbol{\mu}$ and those that do not.

(c) Take the derivative with respect to $\boldsymbol{\mu}$, set to 0, and solve for $\boldsymbol{\mu}$.

First practicals in Machine learning 1 – 2024 – Paper 1

3 Probability theory (September)

Consider the following setting. You are driving down the street at night and suddenly you see a man climbing through a broken window of a jewelry store. Then, he runs away carrying a bag over his shoulder. For many of us, our gut reaction would be to think the man in question is a criminal. Why do we draw this conclusion instead of another scenario? Let's explore this using the methods of Probability Theory.

(a) Explain in words: why would many people draw the conclusion that the man in question is a criminal? Try to think in terms of probability (1-2 sentences are sufficient).

(b) Show, formally (using probability theory), that the probability of us believing the man is a criminal given our observation is based on our beliefs of making this observation when the man is a criminal and making the observation when the man is not a criminal. Define first your variables for the problem and then your answer.

(c) Let's assume one in every 10^5 people is in fact a criminal, the probability of making this observation when the man is not a criminal is $\frac{1}{10^6}$, and that of making this observation when the man is a criminal is 0.8.

Compute the probability of the man being a criminal based on our observations.

(d) The next morning you learn that a group of kids have smashed multiple store fronts in your neighborhood. How does this change your beliefs, i.e., do you still think the man is a criminal? Explicitly state which belief updates you make and re-compute the probability of the man being a criminal given our observation. *Note, you do not have to do a Bayesian update or justify your belief update mathematically.*

First practicals in Machine learning 1 – 2024 – Paper 1

4 MAP of Binomial (September)

Suppose someone gave you a magic coin that even though it is perfectly flat, it shows heads much more than tails. You toss it 8 times, resulting in an 8-bit binary string.

(a) This sequence can be modelled with a binomial distribution. If the coin is unbiased, what will be the value of p?

(b) We observe the bitstring 01110110. What is the maximum likelihood estimate of p? First derive the question analytically, then give the numerical solution.

(c) Knowing this, a naive person would accept the coin as biased. You are not naive, and you've flipped enough coins to know that almost all of them are practically unbiased. You express this by having a strong prior. This prior appears in Bayes' rule. It's a beta distribution, which is the conjugate prior of the binomial, resulting in a beta posterior distribution.

Show that the posterior distribution is beta. Combine all proportionality constants (not depending on p), in a constant Z.

(d) Incorporating your strong beta prior with $a = 13$ and $b = 13$, what is your skeptical (MAP) estimate of p? Again, first derive analytically, then plug in the numbers.

(e) How can a and b be interpreted?

After better inspection, you see that the coin has been made using two different kinds of metal. It seems clear now that this is not a regular coin. How does your belief change? What parameters change, and in which direction?

(f) Could these results have been derived with a Bernoulli distribution, too? If so, explain why and obtain the ML (maximum likelihood) estimate.

5 Probability theory II (September)

For this question, you will compute the expression for the posterior parameter distribution for a simple data problem. Assume we observe N univariate data points $x_1, x_2, ..., x_N$. Furthermore, we assume that they are generated by a Gaussian distribution with known variance σ^2, but unknown mean μ. Assume a prior Gaussian distribution over the unknown mean, i.e., $p(\mu) = \mathcal{N}(\mu|\mu_0, \sigma_0^2)$. When answering these questions, use $\mathcal{N}(a|b, c^2)$ to indicate a Gaussian (normal) distribution over a with mean b and variance c^2.You do not need to write down the explicit form of a Guassian distribution.

(a) Write down the general expression for a posterior distribution, using θ for the parameter, $\mathcal{D}$ for the data. Indicate the prior, likelihood, evidence, and posterior.

(b) Write the posterior for this particular example. You do not need an analytic solution.