Deep Learning (SS2020)
Seminar 5

May 22, 2020

Assignment 1 (Dropout, Bernoulli)

a. Let us see how to design dropout noises more conveniently for implementation. Consider the following Bernoulli noises:

$$Z = \begin{cases} a, & \text{with probability } p \\ 0, & \text{with probability } 1-p \end{cases} \tag{1}$$

What should be the value of a so that $\mathbb{E}[Z] = 1$? This will allow us to avoid rescaling of the weights at the test time and just apply this noise at the training time.

b. Sometimes randomized procedures are used to quantize the gradients (for faster communication in a distributed system). Let $g_i \in \mathbb{R}$ be components of the gradient computed in the worker, $i = 1 \dots n$. Suppose we want to encode the gradient using 1 bit per component and possibly a few real numbers for the whole sequence. Let $\tilde{g}_i$ be the binary encoding. How to chose this encoding in a randomized way so that $\mathbb{E}[\tilde{g}] = g$ and hence we preserve the guarantee of an unbiased (but more noisy) gradient estimate?

Assignment 2 (Ridge Regression)
Consider linear regression model with

$$y_i = w^{\mathsf{T}} x_i + \varepsilon_i, \tag{2}$$

where i is a data point, $x_i \in \mathbb{R}^n$, $y_i \in \mathbb{R}$, $w \in \mathbb{R}^n$ is the weight vector and $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$ are independent measurement errors.

a. Formulate the maximum likelihood learning for this problem and express the log likelihood. You should get the quadratic loss function.
Hint: Since $\varepsilon_i = y_i - w^\mathsf{T} x_i$ is normally distributed, we have $p(y_i|x_i) = p_\mathcal{N}(y_i - w^\mathsf{T} x_i, \sigma^2)$.

b. Consider now also the augmentation with noises in the inputs:

$$y_i = A(x_i + \xi_i) + \varepsilon_i, \tag{3}$$

where $\xi_i \sim \mathcal{N}(0, \lambda^2 I_n)$ are independent. Compute the expected value in ξ of the quadratic loss function derived above. You should obtain a variant of weight decay regularization term.

($\star$) Formulate the maximum likelihood learning and write the log likelihood for the problem (3). What regularization we get in this case?
Hint: write the likelihood of the observed data point (x_i, y_i) integrating out the unobserved noises ξ_i, ε_i.

Assignment 3 Compute the KL-divergence of two univariate normal distributions. Try to generalise your result to multivariate normal distributions.

Assignment 4 Let $X \in \mathbb{R}$ be a normally distributed random variable, i.e.

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{(x-\mu)^2}{2\sigma^2}}.$$

Prove the equality

$$\frac{\partial}{\partial \mu} \mathbb{E}_{\mathcal{N}(\mu,\sigma)} f(x) = \mathbb{E}_{\mathcal{N}(\mu,\sigma)} f'(x),$$

where $f'(x)$ denotes the derivative of f. Hint: use the substitution $\tilde{x} = (x - \mu)/\sigma$ in the integral for the expectation.