Exercise 1: Bias and Variance of Mean Estimators (20 P)

Assume we have an estimator $\hat{\theta}$ for a parameter θ . The bias of the estimator $\hat{\theta}$ is the difference between the true value for the estimator, and its expected value

$$\operatorname{Bias}(\hat{\theta}) = \operatorname{E}[\hat{\theta} - \theta].$$

If $\mathrm{Bias}(\hat{\theta})=0$, then $\hat{\theta}$ is called unbiased. The variance of the estimator $\hat{\theta}$ is the expected square deviation from its expected value

$$\operatorname{Var}(\hat{\theta}) = \operatorname{E}[(\hat{\theta} - \operatorname{E}[\hat{\theta}])^2].$$

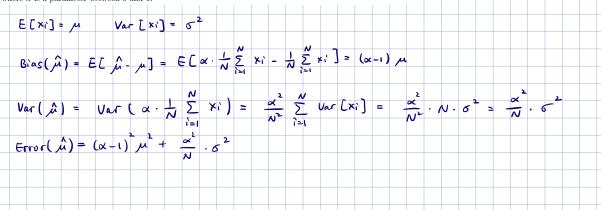
The mean squared error of the estimator $\hat{\theta}$ is

$$\operatorname{Error}(\hat{\theta}) = \operatorname{E}[(\hat{\theta} - \theta)^2] = \operatorname{Bias}(\hat{\theta})^2 + \operatorname{Var}(\hat{\theta})$$

Let X_1,\ldots,X_N be a sample of i.i.d random variables. Assume that X_i has mean μ and variance σ^2 . Calculate the bias, variance and mean squared error of the mean estimator:

$$\hat{\mu} = \alpha \cdot \frac{1}{N} \sum_{i=1}^{N} X_i$$

where α is a parameter between 0 and 1.



Exercise 2: Bias-Variance Decomposition for Classification (30 P)

The bias-variance decomposition usually applies to regression data. In this exercise, we would like to obtain similar decomposition for classification, in particular, when the prediction is given as a probability distribution over C classes. Let $P = [P_1, \dots, P_C]$ be the ground truth class distribution associated to a particular input pattern. Assume a random estimator of class probabilities $\hat{P} = [\hat{P}_1, \dots, \hat{P}_C]$ for the same input pattern. The error function is given by the expected KL-divergence between the ground truth and the estimated probability distribution:

$$Error = E[D_{KL}(P||\hat{P})] = E[\sum_{i=1}^{C} P_i \log(P_i/\hat{P}_i)]$$

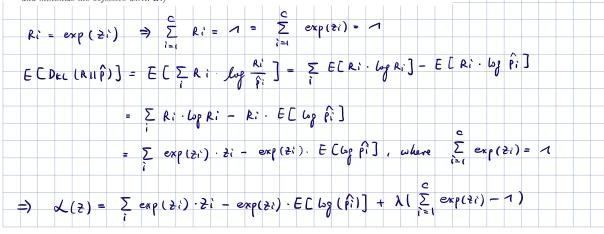
First, we would like to determine the mean of of the class distribution estimator \hat{P} . We define the mean as the distribution that minimizes its expected KL divergence from the class distribution estimator, that is, the distribution R that optimizes

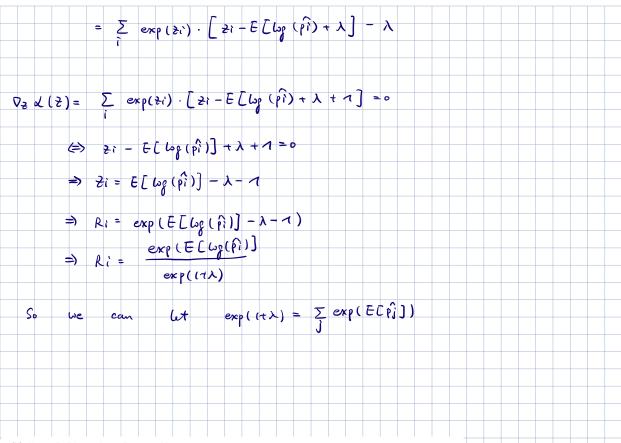
$$\min_{\mathbf{p}} \ \mathrm{E}\big[D_{\mathrm{KL}}(R||\hat{P})\big]$$

(a) Show that the solution to the optimization problem above is given by

$$R = [R_1, \dots, R_C] \quad \text{where} \quad R_i = \frac{\exp \mathbf{E} \big[\log \hat{P}_i\big]}{\sum_j \exp \mathbf{E} \big[\log \hat{P}_j\big]} \qquad \forall \; 1 \leq i \leq C.$$

(Hint: To implement the positivity constraint on R, you can reparameterize its components as $R_i = \exp(Z_i)$, and minimize the objective w.r.l. Z.)





(b) Prove the bias-variance decomposition

$$\operatorname{Error}(\hat{P}) = \operatorname{Bias}(\hat{P}) + \operatorname{Var}(\hat{P})$$

where the error, bias and variance are given by

$$\operatorname{Error}(\hat{P}) = \operatorname{E}\big[D_{\operatorname{KL}}(P||\hat{P})\big], \qquad \operatorname{Bias}(\hat{P}) = D_{\operatorname{KL}}(P||R), \qquad \operatorname{Var}(\hat{P}) = \operatorname{E}\big[D_{\operatorname{KL}}(R||\hat{P})\big].$$

(Hint: as a first step, it can be useful to show that $E[\log R_i - \log \hat{P}_i]$ does not depend on the index i.)

$$E[\log Ri - \log \hat{fi}] = E[\log \frac{\exp(E[\log fi])}{\exp(i\pi\lambda)} - \log \hat{fi}]$$

$$= E[E[\log fi] - (i\pi\lambda) - \log \hat{fi}]$$

$$= -(i\pi\lambda)$$

$$E[\log fi] + var(\hat{f}) = Dec(pir) + E[Dec(Rii \hat{f})]$$

$$= \sum_{i} p_{i} \cdot ln \frac{p_{i}}{R_{i}} + E[\sum_{i} R_{i} \cdot ln \frac{R_{i}}{\hat{f}_{i}}]$$

$$= \sum_{i} p_{i} \cdot ln \frac{p_{i}}{R_{i}} + E[\sum_{i} R_{i} \cdot ln \frac{R_{i}}{\hat{f}_{i}}]$$

$$= \sum_{i} p_{i} \cdot ln \frac{p_{i}}{R_{i}} + E[\sum_{i} R_{i} \cdot ln \frac{R_{i}}{\hat{f}_{i}}]$$

$$= \sum_{i} p_{i} \cdot ln \frac{p_{i}}{R_{i}} + E[\sum_{i} R_{i} \cdot ln \frac{R_{i}}{\hat{f}_{i}}]$$

F / E .	. (2)	
= t(2 r	1 68 Pr - Pr 6	g Ri + Pi 65 Ri - 1	Pilog Pi)	
·				
= €(Σ p	of log pi - pi	; R; + P; 6; R; -		
1 .	0 .	0		
	D ((&))			
= E(VC)	. [4 [7]			
= error (<u>^</u>			
= emi (ry			