KDGs Bæckground Reading

- · What are Polytopes? a geometric object with flat sides.
- · Kernel Density Estimation (KDE)
 - a non-parametric way to estimate the PDF of a random variable.
 - Let $(\alpha_1, \alpha_2, ..., \alpha_n)$ be i.i.d. sampled drawn from a known distribution f at a given point α . We are interested in estimating the shape of the function f.

$$f_h(\alpha) = \frac{1}{h} \sum_{i=1}^{h} K_h(n-\alpha_i) = \frac{1}{hh} \sum_{i=1}^{h} K\left(\frac{\alpha-\alpha_i}{h}\right)$$

where $k_n \rightarrow \text{Kernel } (a \text{ non-negative } f^{\underline{h}})$ $h \rightarrow \text{Bandwith } (a \text{ smoothing parameter})$

· Performance Metrics

1 Expected Caliberation From CECE)

Measures the expected difference between accuracy and confidence by grouping all samples (Size N) in to K bins & calculating

ECE =
$$\sum_{i=1}^{k} \frac{|B_i|}{N} |acc_i - conf_i|$$

acci and confi are accuracy & average confidence in the i-th bin & |Bil is the number of samples in bin Bo

- The pseudo-probabilities are class probabilities we get from the final layer of a NN.

 The pseudo-probability of the predicted class generally over-estimates the actual probability of getting a correct answer.
- If this over-estimation can be measured, it can be used to caliberate the NN such that it's pseudo-probabilities would match the achial probability of the classes.

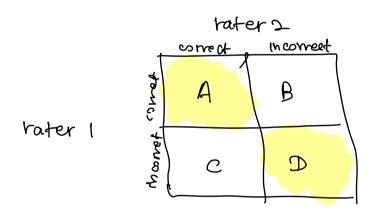
2 Cohen's Kappa

$$\mathcal{K} = \frac{P_0 - Pe}{1 - Pe}$$

Pe -> hypothetical probability of chance agreement.

k Measures the agreement between 2 raters who each classify N Hmes in to C mutually exclusive categories.

for 2 raters that are rating the same thing, corrected for how often that the raters may agree by chance.



In A & D, the two raters are in agreement.

Expected probability that both would say correct => $P(correct) = \frac{A+B}{A+B+C+D} \times \frac{A+C}{A+B+C+D}$

Expeched probability that both would say incorret=>

$$P(\text{incorrect}) = \frac{C+D}{A+B+C+D} \times \frac{B+D}{A+B+C+D}$$

Pe = overall random agreement probability that they agreed on either yes or no.

Then
$$K = \frac{Po - Pe}{1 - Pe}$$

Understanding KDN

& Understanding Polytopes formed by MLPs

- partition & vote
- representation space learnt by DNNs is a partitioning of feature space in to a union of convex polytopes.

classical statistical formulation of classification problem >

$$(x_1, Y_1), (x_2, Y_2), ..., (x_n, Y_n) \stackrel{id}{\sim} f_{xy}$$
 $T_n = \{(x_1, Y_1)\}_{1 \in \{1, ..., n\}}$: training data
 (x, y) : to-be classified test observation & its true label

 $x \in \mathbb{R}^d$ $Y \in \{0,1\}$

Learn a classification rule $g_n = g(\cdot; T_n)$ that maps feature vectors to class labels S.t. probability of misclassification $L(g_n) = P[g(x; T_n) \neq Y | T_n]$ is small.

Stone's The for universally consistent classification:

- a successful classifier can be constructed by
- on n, such that the * training points in each cell goes to infinity but clowly in n
- Formating the parterior $\eta(x) = P[Y=1 | X=x]$ locally by voting based on the training class labels associated with the training feature vectors in cell $C(x) \subset \mathbb{R}^d$ in which the test observation falls.

Then $L(g_n) \to L^*$ almost surely for any F_{xy} where L^* 1s the Bayes optimal probability of mis classification.