

Measuring Classification Decision Certainty and Doubt

Alexander M. Berenbeim¹, Iain J. Cruickshank¹, Susmit Jha², Robert H. Thomson¹, and Nathaniel D. Bastian¹

¹Army Cyber Institute, United States Military Academy, West Point, NY 10996, USA

²Computer Science Laboratory, SRI International, Menlo Park, CA 94025, USA

Abstract

Quantitative characterizations and estimations of uncertainty are of fundamental importance in optimization and decision-making processes. Herein, we propose intuitive scores, which we call *certainty* and *doubt*, that can be used in both a Bayesian and frequentist framework to assess and compare the quality and uncertainty of predictions in (multi-)classification decision machine learning problems.

Introduction

The rising use of artificial intelligence and machine learning technologies to power intelligent systems has led to a growing desire for automating, accelerating, and enhancing decision-making processes in safety-critical applications. These technologies offer decision-makers the ability to gain an information and decision-making advantage at the speed of machines. However, in safety-constrained decision-making, it is crucial to estimate and factor in the level of certainty and doubt associated with each classification decision [3]. In such scenarios, even small chances of risky outcomes may have a significant impact on classification decisions, regardless of the most probable predicted outcome, and safety-critical applications need to be sensitive to such tail probabilities [1], [2].

Furthermore, in the (multi-class) classification decision setting when model-assigned probabilities are close to uniformly distributed, we have cause to doubt the model's prediction, even if it is accurate. Moreover, from a theoretical view of probability, we have cause to doubt the model architecture if the greatest predicted probability is arbitrarily close to the second greatest probability. An intuitive score that can capture this sense of certainty and doubt about our predictions is desirable.

Defining Certainty and Doubt

Working in the context of (multi-class) classification decision problems on N distinct labels, we will assume that we have a learning machine $f(\omega, \mathbf{x})$ that on input $\mathbf{x} \in \mathbb{R}^{Input}$ with weights ω , outputs a probability vector $\mathbf{p} \in \Delta^N$ after applying a differentiable function to normalize a vector $\mathbf{y} \in \mathbb{R}^N$, such as SOFTMAX, denoted by σ .

Let $j = \arg \max_i y_i$ and $\hat{p} = \max_i \pi_i(\mathbf{p})$, where π_i are the canonical projection maps onto the i^{th} component.

We define *certainty* and *doubt* as reciprocals, the geometric intuition behind this will be described in the next section. We define the *pairwise certainty* and *pairwise doubt* by

$$\chi_i = \begin{cases} 1 & i = j \\ \pi_j(\mathbf{p}) - \pi_i(\mathbf{p}) \equiv (\pi_j - \pi_i)(\mathbf{p}) & i \neq j \end{cases} \quad (\text{Pairwise Certainty})$$

and

$$\delta_i = \begin{cases} 0 & i = j \\ \frac{1}{\pi_j(\mathbf{p}) - \pi_i(\mathbf{p})} & i \neq j \end{cases} \quad (\text{Pairwise Doubt})$$

The intuitive motivation for these definitions is that if f gives two components the same maximum score, then the model is ‘maximally’ uncertain between the two labels, whence the certainty is 0 and doubt is infinite, while there should be no doubt when comparing one label score with itself; in order to reasonably doubt, an alternative must be possible. Our pairwise definitions can be used to define vector-valued maps from the probability simplex Δ^N to \mathbb{R}^N and $\hat{\mathbb{R}}^N := \mathbb{R}^N \cup \{\infty\}$ respectively via $\chi = \chi_1 \times \cdots \times \chi_N$ and $\delta = \delta_1 \times \cdots \times \delta_N$ for certainty and doubt, respectively. Moreover, the non-diagonal components of certainty and doubt are distributed between $[0, \hat{p}]$ and $[\hat{p}^{-1}, \infty]$.

Further, for the ease of graphing the distribution of doubt scores, we may also consider either using $-\log$ post-composed with the pairwise certainty scores or the difference between the log-probabilities as a related notion of doubt. This latter definition can be straight forwardly computed without using log-softmax by computing $\hat{y}\mathbf{1} - \mathbf{y}$ before applying softmax, although care will need to be taken with respect to the index of maximal value. We define certainty and doubt in this setting as *raw certainty* and *raw doubt*. Formally:

$$\xi_i = \begin{cases} 1 & i = j \\ \pi_j(\mathbf{y}) - \pi_i(\mathbf{y}) \equiv (\pi_j - \pi_i)(\mathbf{y}) & i \neq j \end{cases} \quad (\text{Pairwise Raw Certainty})$$

$$\rho_i = \begin{cases} 0 & i = j \\ \frac{1}{\pi_j(\mathbf{y}) - \pi_i(\mathbf{y})} & i \neq j \end{cases} \quad (\text{Pairwise Raw Doubt})$$

Finally, although it is more computationally expensive, we can also encode certainty across the assigned probability vector \mathbf{p} with the following skew-symmetric matrix:

$$\mathbf{C}^o(\mathbf{p}) := \mathbf{p}\mathbf{1}^T - \mathbf{1}\mathbf{p}^T$$

and set certainty as

$$\mathbf{C}(\mathbf{p}) = \mathbf{I} + \mathbf{C}^o(\mathbf{p}).$$

Since $\mathbf{C}(\mathbf{p})$ will always be invertible, we may naively suppose that \mathbf{C}^{-1} is a good candidate for expressing doubt. However, consider the following binary classification vector $\mathbf{p} = \begin{bmatrix} .5 \\ .5 \end{bmatrix}$. Then $\mathbf{C} = \mathbf{I}$ and $\mathbf{C}^{-1} = \mathbf{I}$, while intuitively we would be maximally uncertain between the two labels as the probability vector indicates they have equal likelihood. Instead, we may find doubt may be better described as

$$\mathbf{D}(\mathbf{p}) = \text{Inv}(\mathbf{C}(\mathbf{p})) - \mathbf{I}$$

where Inv is the element-wise inverse function.

We propose developing scores from $\mathbf{C}(\mathbf{p})$ and $\mathbf{C}^o(\mathbf{p})$ for the purposes of quantifying uncertainty between different neural network architectures. For the ease of graphing the distribution of doubt and for studying the statistics of the doubt scores between models, we suggest setting our minimum certainty/maximum doubt score to be set to

$$\min_i \pi_i(\mathbf{x}) : \max_{\mathbf{x} \in \text{Row}(\mathbf{C}(\mathbf{p}))} \|\mathbf{x}\|_1 \quad \text{or equivalently} \quad = \|\mathbf{D}(\mathbf{p})\|'_\infty,$$

where $\|\mathbf{D}\|'_\infty$ is the extended Chebyshev metric adding a point at infinity, with the goal of maximizing minimum certainty, or equivalently, minimizing maximum doubt.

Geometric Interpretation

Projective Geometry

A *projective space* over a field K is a set of one-dimensional subspaces of the vector space K^{n+1} , which we will denote by $\mathbb{P}^n(K)$. This can be equivalently understood as the quotient of $K^{n+1} \setminus \{\mathbf{0}\}$ by the action of K^\times acting by scalar multiplication, i.e. $\mathbb{P}^n(K)$ can be identified as a set of equivalence classes $[\mathbf{p}]$ such that any two non-zero points \mathbf{p}, \mathbf{q} in $\mathbb{P}^n(K)$ belong to the same equivalence class if $\mathbf{p} = c\mathbf{q}$ for some $c \in K^\times$. Points in projective vector spaces are usually written as *homogeneous vectors* $[Z_0 : \cdots : Z_N]$.

We are primarily interested in the geometry of *real projective* spaces, which are typically denoted by \mathbb{RP}^n . Real projective spaces are well-studied examples of real manifolds, with local coordinates U_k of \mathbb{RP}^n given by

$$\begin{aligned} U_0 &= \{[1 : x_1 : x_2 : \cdots : x_N] \mid x_1, x_2, \dots, x_N \in \mathbb{R}\} \\ U_1 &= \{[x_0 : 1 : x_2 : \cdots : x_N] \mid x_1, x_2, \dots, x_N \in \mathbb{R}\} \\ &\vdots \\ U_N &= \{[x_0 : x_1 : x_2 : \cdots : 1] \mid x_1, x_2, \dots, x_N \in \mathbb{R}\}. \end{aligned}$$

It is readily seen that each U_i is analytically isomorphic to \mathbb{R}^N . Moreover,

$$\mathbb{RP}^N \cong \mathbb{R}^N \cup \mathbb{R}^{N-1} \cup \dots \cup \mathbb{R} \cup \mathbb{R}^0$$

which follows from the identification of

$$\mathbb{RP}^N = U_0 \cup \{[0 : x_1 : \cdots : x_N] \mid [x_1 : \cdots : x_N] \in \mathbb{RP}^{N-1}\} \cong \mathbb{R}^N \cup \mathbb{RP}^{N-1}.$$

In particular, $\{[0 : 0 : \cdots : 0 : 1]\} \cong \mathbb{R}^0$ is referred to as *the point at infinity*.

Since spheres $S^N \cong \mathbb{RP}^N$ by quotienting out by the antipodal points, and the standard Riemannian metric on S^N is invariant under the action of the group $\mathbb{Z}/(2)$, we can push down the metric on S^N to \mathbb{RP}^N . Explicitly, the metric on \mathbb{RP}^N is defined by

$$\langle v, w \rangle_{[p]} = \langle (dq|_p)^{-1}(v), (dq|_p)^{-1}(w) \rangle_p,$$

with $q : S^N \rightarrow S^N / \sim_\alpha \cong \mathbb{RP}^N$ the quotient map of S^N by the antipodal equivalence relation \sim_α and $[p] = \{\pm p\}$ for $p \in S^N$. Further, this is well defined since the antipodal map $\alpha : S^N \rightarrow S^N$ is an isometry, so $q = q \circ \alpha$, and $d\alpha_p$ will also be an isometry.

The Geometry of Doubt

Certainty and doubt can be naturally viewed using projective geometry.

Consider the case where we have N labels, and Δ^N is the probability simplex, e.g.

$$\Delta^N = \{\mathbf{p} \in \mathbb{R}^N \mid \mathbf{1}_N \cdot \mathbf{p} = 1 \wedge \bigwedge_{1 \leq i \leq N} \pi_i(\mathbf{p}) \geq 0\},$$

where $\pi_i : \mathbb{R}^N \rightarrow \mathbb{R}$ is the canonical i^{th} -coordinate projection function.

Further, we may always permute the indices as necessary so that \mathbf{p} is such that $\pi_i(\mathbf{p}) \geq \pi_{i+1}(\mathbf{p})$ for $i \in [N]_+$. Denote Δ_{\geq}^+ to be this ordered simplex.

We let $F : \Delta_{\geq}^N \rightarrow \prod_{i \in [N]_+} \mathbb{RP}^1$ be given coordinate wise by

$$f_i(\mathbf{p}) = \begin{cases} [1 : 0] & i = 1 \\ [\pi_1(\mathbf{p}) - \pi_i(\mathbf{p}) : 1] & i \neq 1 \end{cases}$$

Since the first coordinate we're mapping is the distance between p_1 and p_i , this can be understood as expressing our 'certainty' that the first label is correct over label i , while the second coordinate can be thought of as expressing our doubt. In particular, we have absolute certainty when comparing one probability with itself, as there is no point of comparison. Moreover, since $[p_1 - p_i : 1]$ is a point in an equivalence class, we see that $[p_1 - p_i : 1] = [1 : \frac{1}{p_1 - p_i}]$ whenever $p_1 > p_i$, so that pairwise doubt and certainty identify the same point on the real projective line, while whenever $p_1 = p_i$, $f(p_i) = [0 : 1]$, i.e. we have a point at infinity, indicating we have infinite doubt and zero certainty that our top choice is in fact the correct one.

Further, because the image of Δ_{\geq}^N in $(\mathbb{RP}^1)^N$ under F is identical to $\{[1 : 0]\} \times (\mathbb{RP}^1 \setminus \{[1 : 0]\})^{N-1}$, the image will be isomorphic with $(\mathbb{RP}^1 \setminus \{[1 : 0]\})^{N-1}$. Thus we can consider $F : \Delta_{\geq}^N \rightarrow \prod_{i=2}^N \mathbb{RP}^1$ or alternately,

$F' : \Delta_{\geq}^N \rightarrow \prod_{i=2}^N S^1$ instead.

Proposed Cost Function With Certainty and Doubt Scoring

Taking the projective view, doubt-minimization is a program that aims to avoid estimated probability assignments whose Segre embedding is a degenerate point lying in a real-projective space. In particular, we wish to avoid $\prod_{i=2}^N (p_1 - p_i) = 0$, which indicates that we have no confidence between our assigned label and another label choice. On the other hand, there is no intrinsic reason that we should prefer a non-uniform distribution on the other assigned probabilities below the maximum estimated probability.

When computing a cost function that invokes the doubt, or certainty of a score, we wish to have some appropriate smooth function. We propose using the stereographic projection diffeomorphism $F : S^1 \rightarrow \mathbb{RP}$ given by sending $\theta \mapsto \begin{cases} [1 - \sin \theta : \cos \theta] & \theta \neq \frac{\pi}{2} \\ [\cos \theta : 1 + \sin \theta] & \theta \neq -\frac{\pi}{2} \end{cases}$, with inverse $G : \mathbb{RP} \rightarrow S^1$ given by

$$[a : b] \mapsto \begin{cases} \arcsin \left(\frac{b^2 - a^2}{a^2 + b^2} \right) & a \neq 0 \\ \frac{\pi}{2} & [a : b] = [0 : 1] \end{cases}.$$

Two natural candidates for cost per sample k are the doubt-cost $\theta_{cost}(\mathbf{p}^{(k)}) = \arcsin \left(\frac{1 - \prod_{i \neq j} \chi_i^2}{1 + \prod_{i \neq j} \chi_i^2} \right)$, and the

raw-doubt cost $\theta_{cost}(\mathbf{y}^{(k)}) = \arcsin \left(\frac{1 - \prod_{i \neq j} \xi_i^2}{1 + \prod_{i \neq j} \xi_i^2} \right)$. The cost functions incorporating doubt should be reflective of the underlying function/phenomenon that f is trying to approximate, and penalize doubt accordingly.

Acknowledgements

This work is supported in part by the U.S. Army Combat Capabilities Development Command (DEVCOM) Army Research Laboratory under Support Agreement No. USMA 21050, as well as the Defense Advanced Research Projects Agency (DARPA) under Support Agreement No. USMA 23004. The views expressed in this paper are those of the authors and do not reflect the official policy or position of the United States Military Academy, the United States Army, the Department of Defense, or the United States Government.

References

- [1] J. Vaicenavicius, D. Widmann, C. Andersson, F. Lindsten, J. Roll, and T. Schön, “Evaluating model calibration in classification,” in *Proceedings of the Twenty-Second International Conference on Artificial Intelligence and Statistics*, K. Chaudhuri and M. Sugiyama, Eds., ser. Proceedings of Machine Learning Research, vol. 89, PMLR, Apr. 2019, pp. 3459–3467. [Online]. Available: <https://proceedings.mlr.press/v89/vaicenavicius>
- [2] A. D. Cobb, B. Jalaian, N. D. Bastian, and S. Russell, “Toward safe decision-making via uncertainty quantification in machine learning,” in *Systems Engineering and Artificial Intelligence*, W. F. Lawless, R. Mittu, D. A. Sofge, T. Shortell, and T. A. McDermott, Eds. Cham: Springer International Publishing, 2021, pp. 379–399, ISBN: 978-3-030-77283-3. DOI: 10.1007/978-3-030-77283-3_19. [Online]. Available: https://doi.org/10.1007/978-3-030-77283-3_19.
- [3] T. Abdelzaher, N. D. Bastian, S. Jha, L. Kaplan, M. Srivastava, and V. V. Veeravalli, “Context-aware collaborative neuro-symbolic inference in iobts,” in *MILCOM 2022 - 2022 IEEE Military Communications Conference (MILCOM)*, 2022, pp. 1053–1058. DOI: 10.1109/MILCOM55135.2022.10017607.