ontotext
making sense of text and data

Maximum Likelihood Estimation with Deep Learning for Multiple Sclerosis Progression Prediction

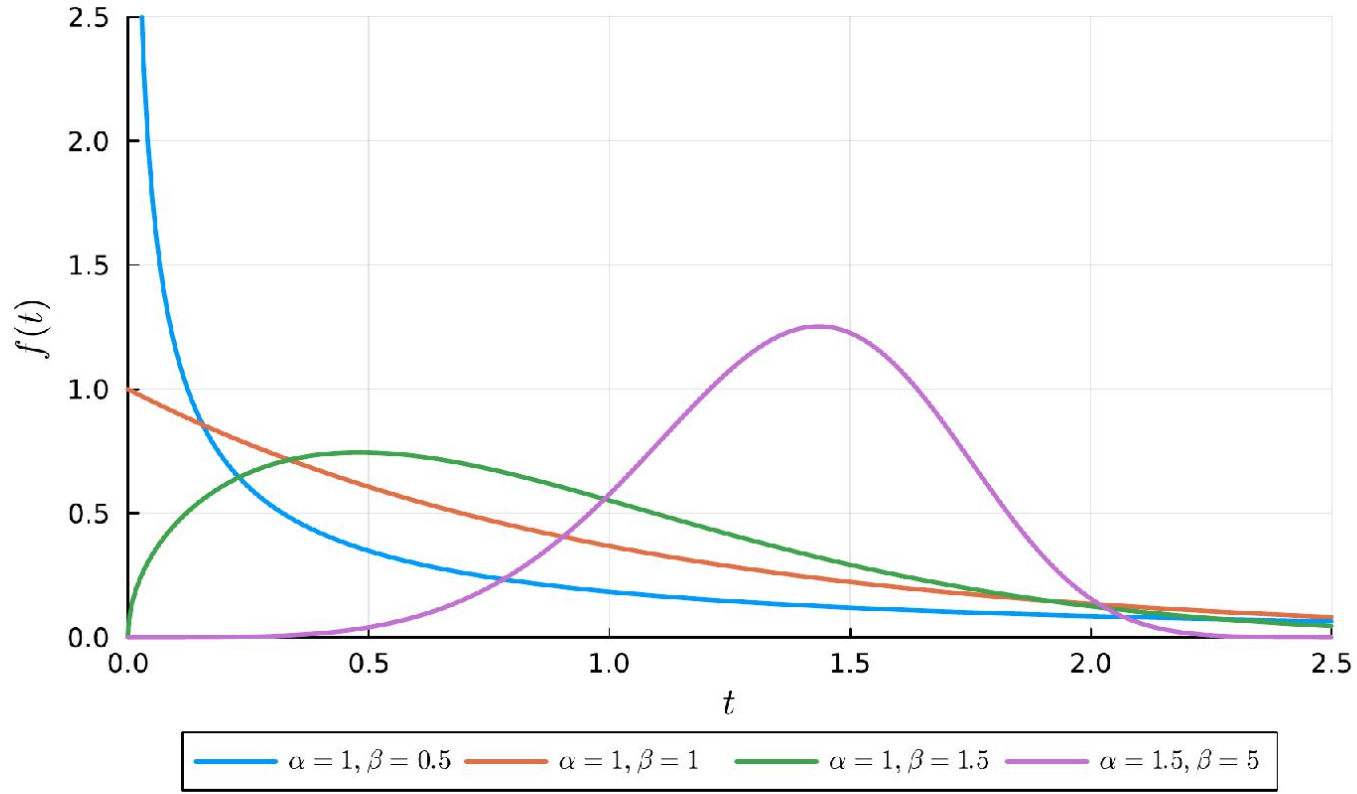
Tsvetan Asamov, Petar Ivanov, Anna Aksenova, Dimitar Taskov, Svetla Boytcheva

September 2023

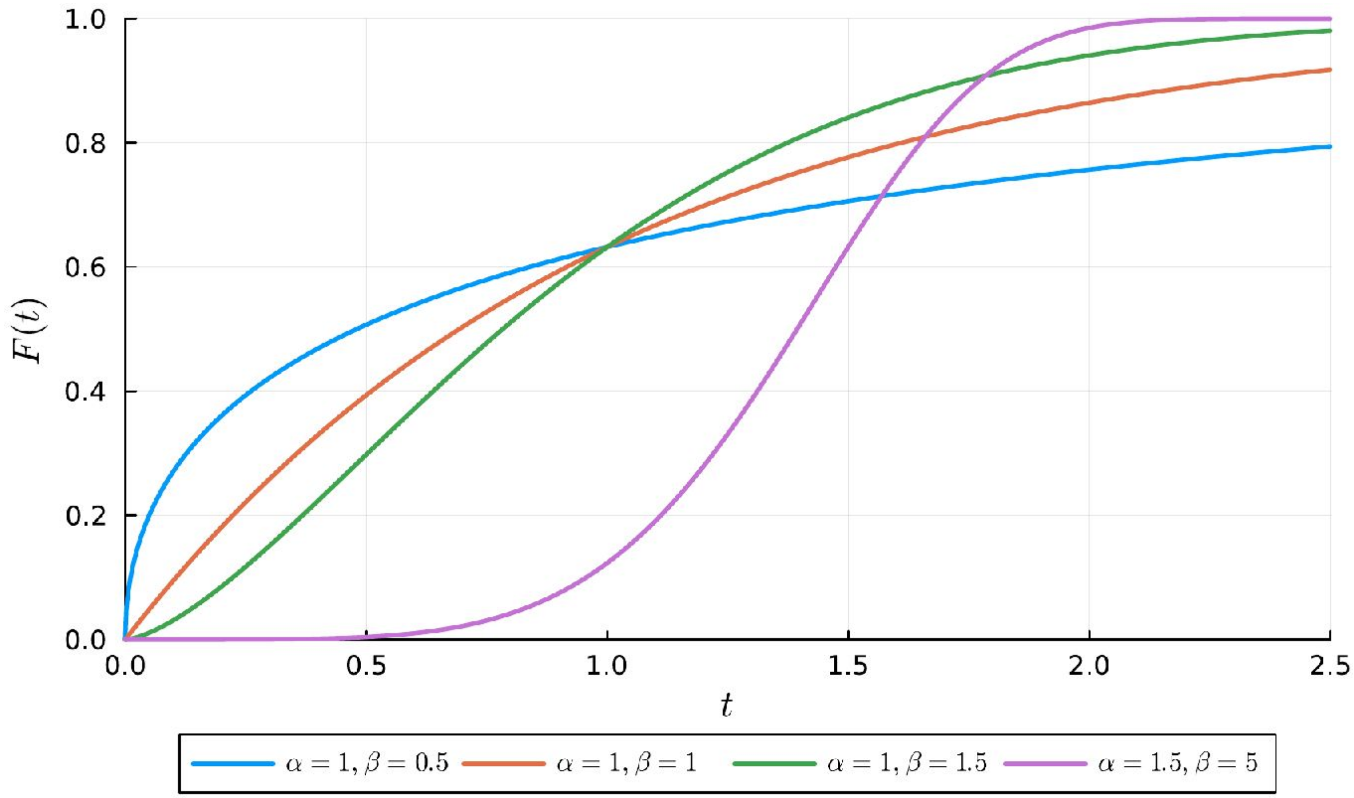
Challenge

- **The iDPP challenge includes the following two tasks:**
  - **Task 1: Predicting the risk of worsening and ranking subjects based on the risk scores.** More specifically, the risk of worsening should be a value between 0 and 1 that reflects how early a patient experiences the worsening event.
  - **Task 2: Predicting the cumulative probability of worsening -** assigning cumulative probability of worsening at different time windows, i.e. between years 0 and 2, 0 and 4, 0 and 6, 0 and 8, 0 and 10.

Weibull Distribution Probability Density Function



Weibull Distribution Cumulative Density Function



ontotext

Weibull Distribution

$$f(t|\alpha,\beta) := \begin{cases} 0, & t < 0 \\ \frac{\beta}{\alpha}\left(\frac{t}{\alpha}\right)^{\beta-1} \exp\left(-\left(\frac{t}{\alpha}\right)^{\beta}\right), & t \geq 0 \end{cases}$$

$$F(t|\alpha,\beta) := \begin{cases} 0, & t < 0 \\ 1 - \exp\left(-\left(\frac{t}{\alpha}\right)^{\beta}\right), & t \geq 0 \end{cases}$$

Likelihood Function

- Assuming independence among patients, we can write the likelihood function as follows:

$$L(\theta) = \prod_{i,\delta_i=1} f(t_i|\theta_i) \prod_{i,\delta_i=0} (1 - F(t_i|\theta_i))$$

Maximum Likelihood Formulation

$$\max_{\mathbf{A},\mathbf{b}} \prod_{i,\delta_i=1} f(t_i|\alpha_i,\beta_i) \prod_{i,\delta_i=0} (1 - F(t_i|\alpha_i,\beta_i))$$

s.t.

$$(\alpha_i,\beta_i) = \Psi(\mathbf{A},\mathbf{b},\mathbf{x}_i),\ i = 1,\dots,I$$

$$\alpha_i > 0,\ i = 1,\dots,I$$

$$\beta_i > 0,\ i = 1,\dots,I$$

where

$$f(t|\alpha,\beta) := \begin{cases} 0, & t < 0 \\ \frac{\beta}{\alpha}\left(\frac{t}{\alpha}\right)^{\beta-1} \exp\left(-\left(\frac{t}{\alpha}\right)^{\beta}\right), & t \geq 0 \end{cases}$$

$$F(t|\alpha,\beta) := \begin{cases} 0, & t < 0 \\ 1 - \exp\left(-\left(\frac{t}{\alpha}\right)^{\beta}\right), & t \geq 0 \end{cases}$$

$$\Psi(\mathbf{A},\mathbf{b},\mathbf{x}) := \sigma(\mathbf{A}_n\sigma(\mathbf{A}_{n-1}\sigma(\dots\mathbf{A}_3\sigma(\mathbf{A}_2\sigma(\mathbf{A}_1\mathbf{x}+\mathbf{b}_1)+\mathbf{b}_2)+\mathbf{b}_3\dots)+\mathbf{b}_{n-1})+\mathbf{b}_n)$$

Maximum Log–Likelihood Formulation

$$\max_{\mathbf{A},\mathbf{b}} \sum_{i,\delta_i=1} \log(f(t_i|\alpha_i,\beta_i)) + \sum_{i,\delta_i=0} \log(1 - F(t_i|\alpha_i,\beta_i))$$

s.t.

$$(\alpha_i, \beta_i) = \Psi(\mathbf{A}, \mathbf{b}, \mathbf{x}_i),\ i = 1, \dots, I$$

$$\alpha_i > 0,\ i = 1, \dots, I$$

$$\beta_i > 0,\ i = 1, \dots, I$$

where

$$f_{\alpha,\beta}(t) := \begin{cases} 0, & t < 0 \\ \frac{\beta}{\alpha}\left(\frac{t}{\alpha}\right)^{\beta-1} \exp\left(-\left(\frac{t}{\alpha}\right)^{\beta}\right), & t \geq 0 \end{cases}$$

$$F_{\alpha,\beta}(t) := \begin{cases} 0, & t < 0 \\ 1 - \exp\left(-\left(\frac{t}{\alpha}\right)^{\beta}\right), & t \geq 0 \end{cases}$$

$$\Psi(\mathbf{A}, \mathbf{b}, \mathbf{x}) := \sigma(\mathbf{A}_n\sigma(\mathbf{A}_{n-1}\sigma(\dots\mathbf{A}_3\sigma(\mathbf{A}_2\sigma(\mathbf{A}_1\mathbf{x} + \mathbf{b}_1) + \mathbf{b}_2) + \mathbf{b}_3 \dots) + \mathbf{b}_{n-1}) + \mathbf{b}_n)$$

Results

- Task 1: Harrell's Concordance Index in (0.6, 0.65) range.
- Task 2: AUROC exceeding 0.8.

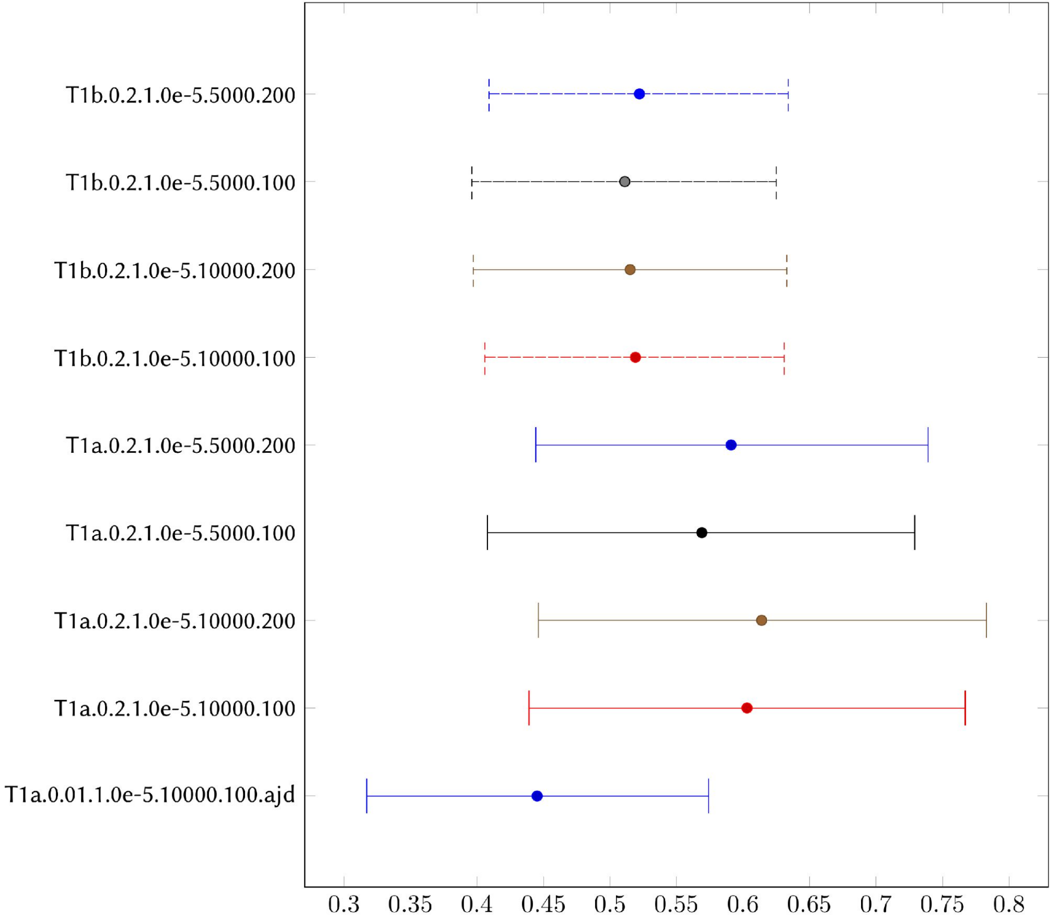


Figure 1: Task 1 - Harrell's Concordance Index computed for all submitted runs. The bars in the plot show the 95% confidence interval.

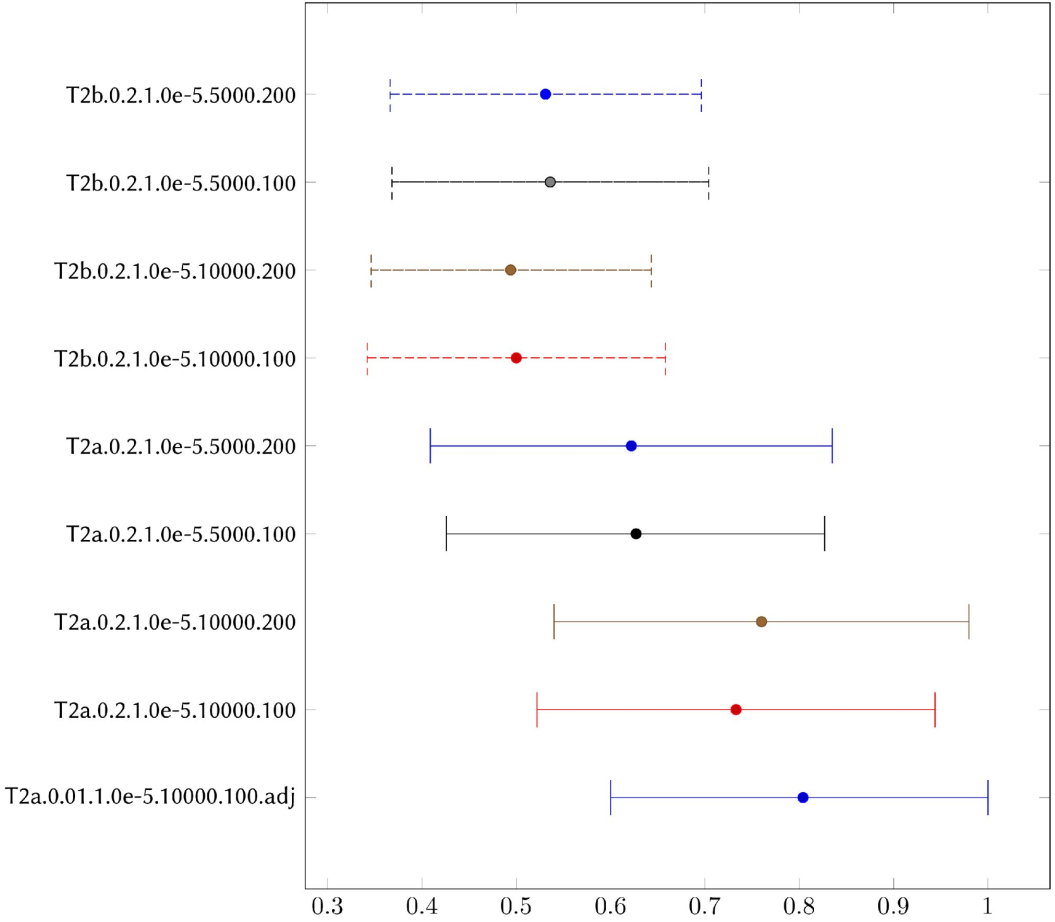
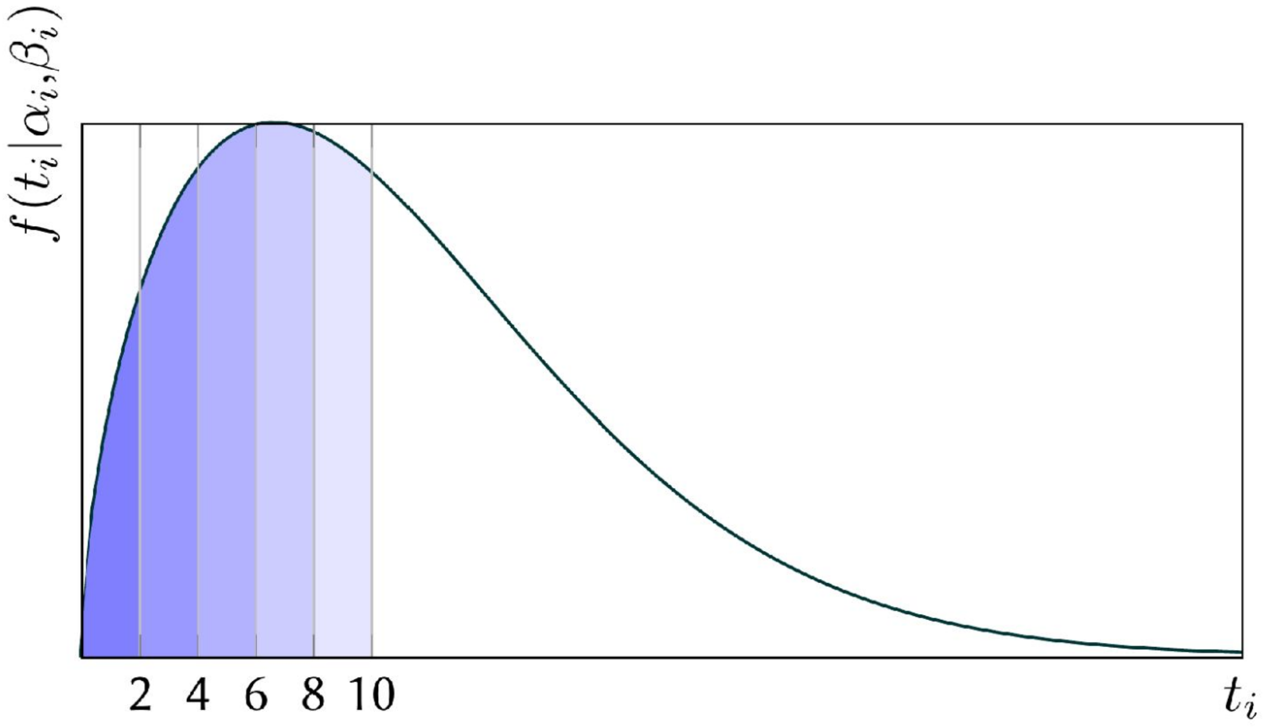
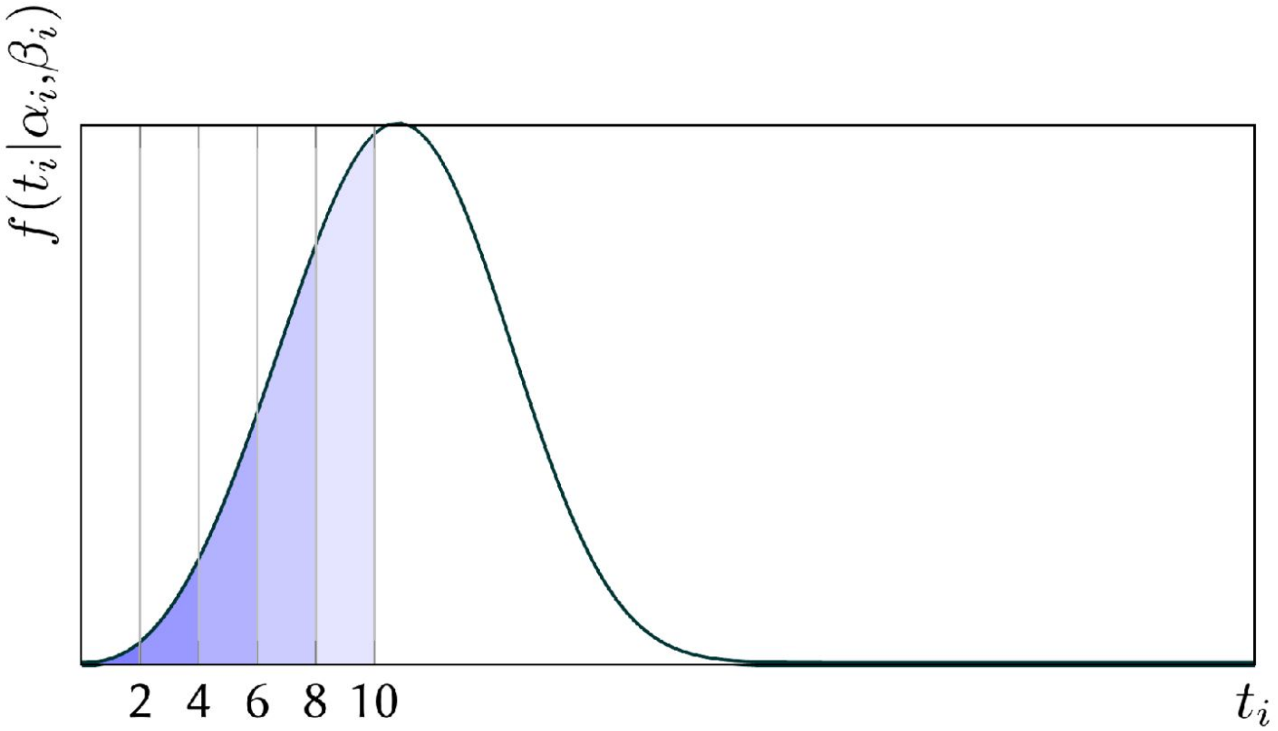


Figure 3: Task 2 - AUROC computed for all submitted runs with a 4-year time horizon. The bars in the plot show the 95% confidence interval.

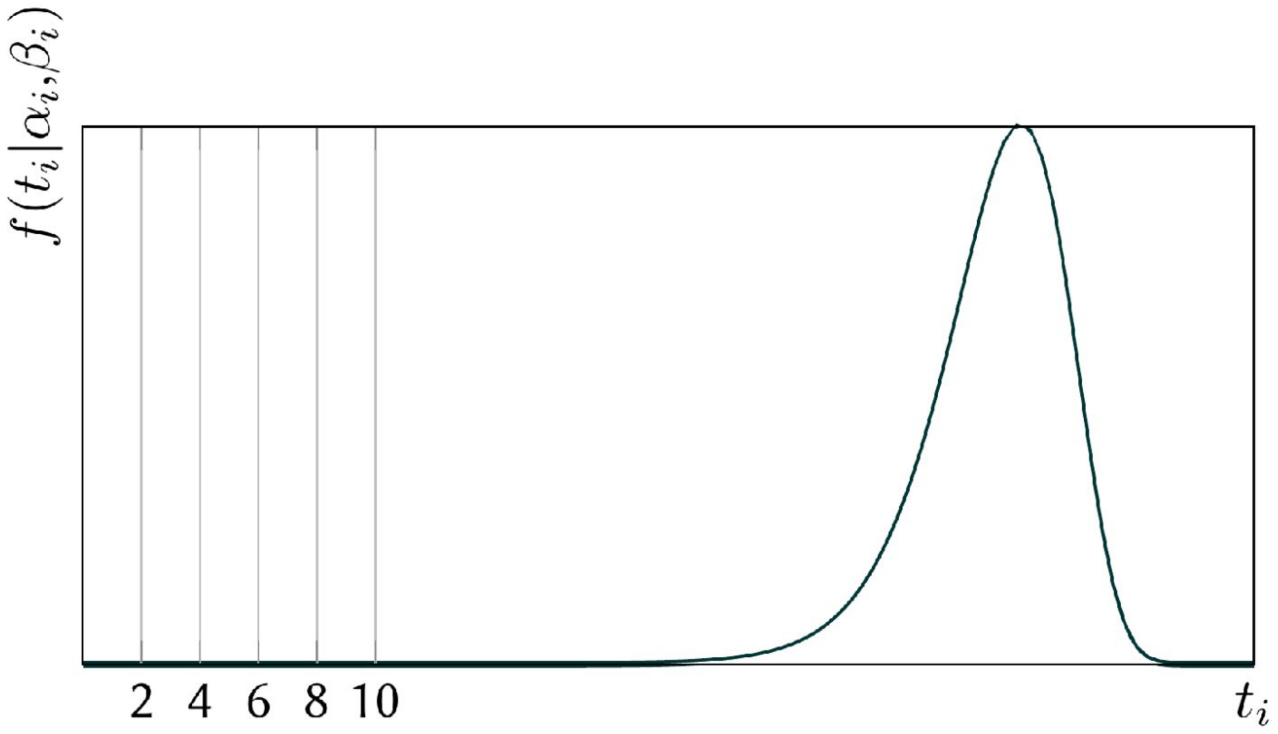
Results - High Risk Probability Density Function

α_i	β_i	Probability Density Function
12.4033	1.58571	

Results - Medium Risk Probability Density Function

11.9824	3.3673	

Results - Low Risk Probability Density Function

32.1712	16.4169	

Discussion

Advantages

- The approach combines the power of maximum likelihood estimation and deep learning.
- The method can be combined with coherent risk measures in order to estimate the risk of worsening.

Limitations

- Lack of convergence and optimality guarantees.
- Currently, our solution does not model mixtures of distributions.
- Numerical results can be further improved.

Conclusion & Further Improvement

- The development of predictive models of the disease is a step forward towards better clinical assessment and an individualized therapeutic approach for multiple sclerosis patients.
- We would like to further improve the quality of the numerical solution with the use of second-order optimization methods such as K-FAC or L-BFGS.
- We can apply scaled-down coherent risk measures in order to obtain risk estimates in the [0, 1] interval.

SMOOTH DATA INTEGRATION

Thank you! Questions?

For more information please see our paper or email us at tsvetan.asamov@ontotext.com

ontotext

References

1. Asamov, T., Aksenova, A., Ivanov, P., Boytcheva, S., & Taskov, D. (2023). Maximum likelihood estimation with deep learning for multiple sclerosis progression prediction. In CLEF.
2. e della Scienza, C. D. S. (2023). Overview of iDPP@ CLEF 2023: The Intelligent Disease Progression Prediction Challenge.
3. Wikimedia Foundation. (2023, August 25). Weibull distribution. Wikipedia. https://en.wikipedia.org/wiki/Weibull_distribution
4. Nagpal, C., Li, X., & Dubrawski, A. (2021). Deep survival machines: Fully parametric survival regression and representation learning for censored data with competing risks. IEEE Journal of Biomedical and Health Informatics, 25(8), 3163-3175.
5. Nagpal, C., Jeanselme, V., & Dubrawski, A. (2021, May). Deep parametric time-to-event regression with time-varying covariates. In Survival Prediction-Algorithms, Challenges and Applications (pp. 184-193). PMLR.