UQ for Credit Risk Management

Deep Evidence Regression approach for Loss Given Default

Paper Presentation, CSE 8803-IUQ

Ashish Dhiman, Anshit Verma

Georgia Tech

Agenda

- **Introduction**
  - Credit Risk Management and Loss Given Default
  - UQ for Loss Given Default
  - Evidential learning
- **Approach**
  - Deep Evidence Regression
    - Pros
    - Limitations
  - Improved Deep Evidence Regression
    - Intuition
    - Improvement made
- **Results**
  - Simulated data
  - Real Data
- **Limitations**

What is Credit Risk ?

- Credit Risk Measures:
  - ***Probability of Default***: Creditor unable to comply with his commitments
  - ***Loss given Default***: Relative fraction of the outstanding amount is lost
  - ***Exposure to Default***: Outstanding amt at the time of default

- Credit Risk Instruments:
  - Fixed Income assets like Corporate Bonds, sovereign bonds from specific countries
  - ***Loans***
  - Mortgages etc.

*M Abdar et al (2021)

UQ for Credit Risk Management?

- Credit Default is a very rare event
  - Data on credit default is very limited
  - In such cases UQ is important because we are not sure of our prediction
  - Quantifying Uncertainty also helps with regulatory angle

- Quantifying uncertainty of machine learning methods for LGD *
  - Deep evidential regression as a UQ framework
  - Bond loss given defaults using Moody's Default and Recovery Database *(not public)*
  - Article also analyses the ratio between aleatoric and epistemic uncertainty
  - Also invokes XAI techniques like ALE plots for regulatory requirements

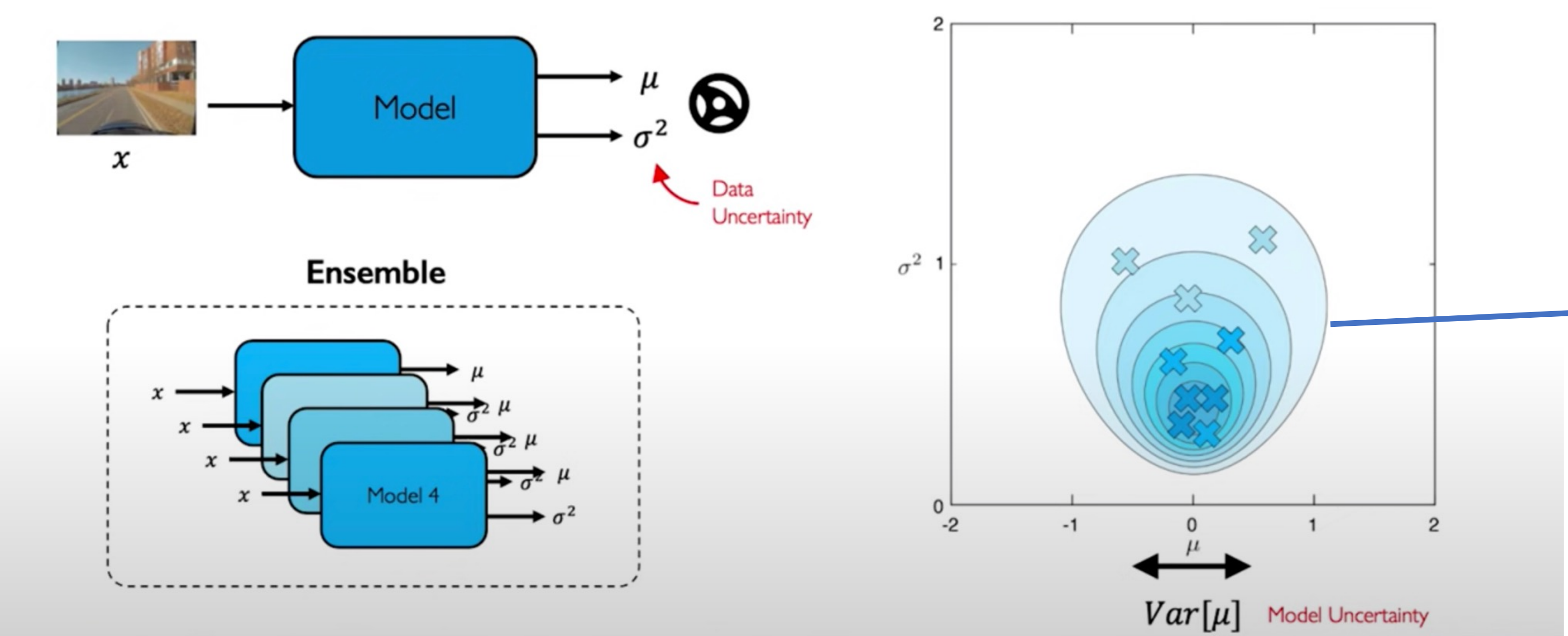
Max Nagl et al (2022)

Georgia Tech

Deep Evidential learning

Model parameters as generated from higher order evidential distribution

- Deep Evidence Classification
- Deep Evidence Regression



Higher Order
Evidential Distribution

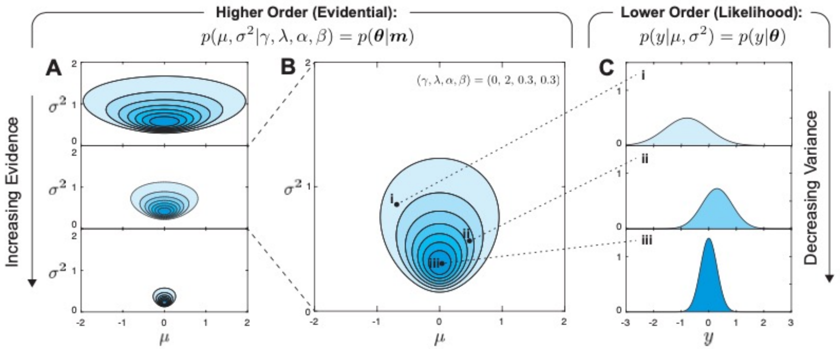
*image source. | Amini et al (NeurIPS 2020)

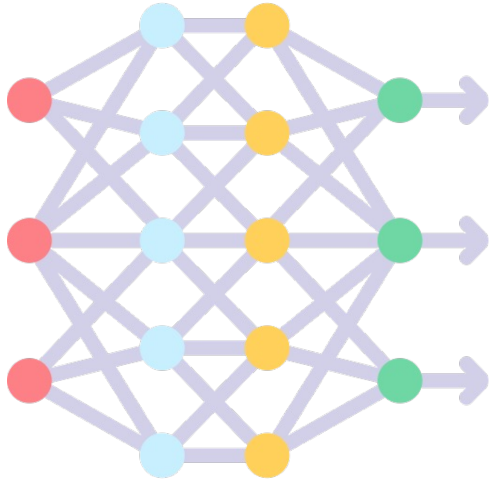
Deep Evidence Regression

- Assumes target variable (y) is Normally distributed with unknown parameters
- Different Prior distributions are then placed on these unknow parameters

$$(y_1, \ldots, y_N) \sim \mathcal{N}(\mu, \sigma^2)$$

$$\mu \sim \mathcal{N}(\gamma, \sigma^2 \upsilon^{-1}) \qquad \sigma^2 \sim \Gamma^{-1}(\alpha, \beta)$$





NN output = m = $\{\alpha, \beta, \nu, \gamma\}$

NN Loss = $L_{\{NLL\}} + c * L_{\{regularisation\}}$

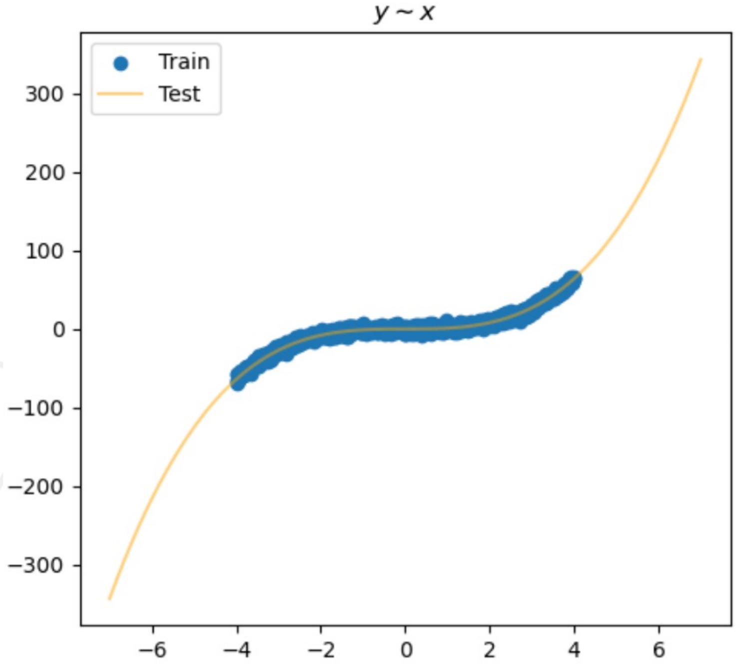
$$p(y_i|\boldsymbol{m}) = \frac{p(y_i|\boldsymbol{\theta}, \boldsymbol{m})p(\boldsymbol{\theta}|\boldsymbol{m})}{p(\boldsymbol{\theta}|y_i, \boldsymbol{m})} = \int_{\sigma^2=0}^{\infty} \int_{\mu=-\infty}^{\infty} p(y_i|\mu, \sigma^2)p(\mu, \sigma^2|\boldsymbol{m})\, \mathrm{d}\mu\, \mathrm{d}\sigma^2$$

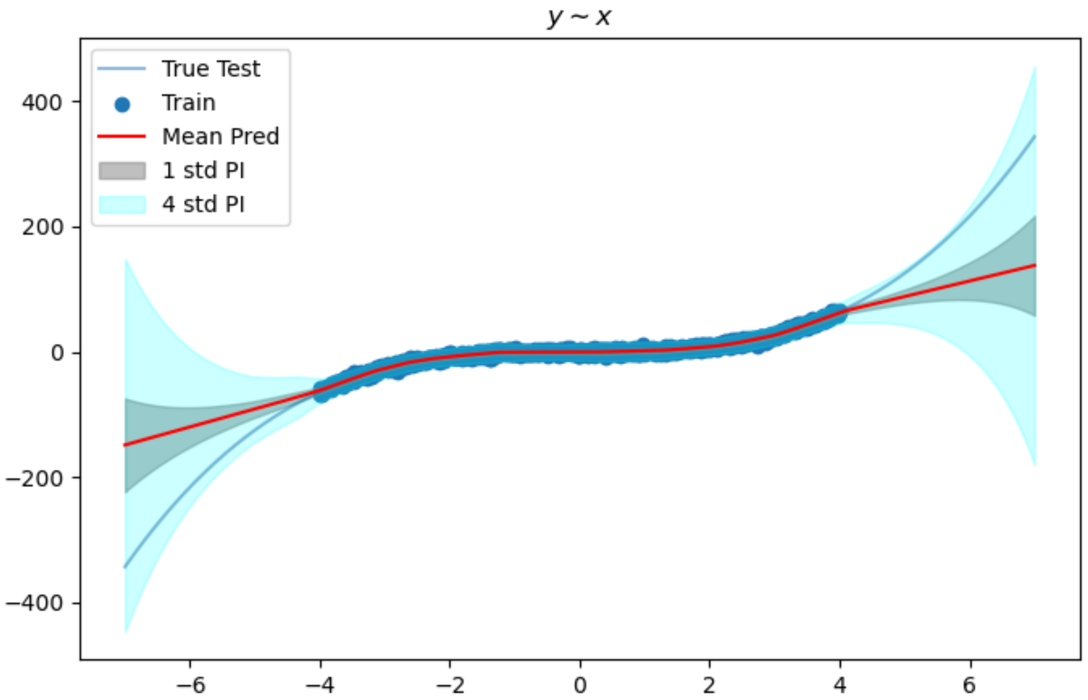
$$\underbrace{\mathbb{E}[\mu] = \gamma}_{\text{prediction}}, \qquad \underbrace{\mathbb{E}[\sigma^2] = \tfrac{\beta}{\alpha - 1}}_{\text{aleatoric}}, \qquad \underbrace{\operatorname{Var}[\mu] = \tfrac{\beta}{\upsilon(\alpha - 1)}}_{\text{epistemic}}.$$

[*M Abdar et al (2021)]()

Deep Evidence Regression: Pros

- Allows analytical computation of uncertainty for OOD Test data
- Example:
  - Data generated by $y = x^3 + \epsilon, \epsilon \sim N\,(0, \sigma^2)$
  - Train set, $x \in [-4,4]$, Test set, $x \in [-7,7]$





*example recreated from Amini et al (NeurIPS 2020)

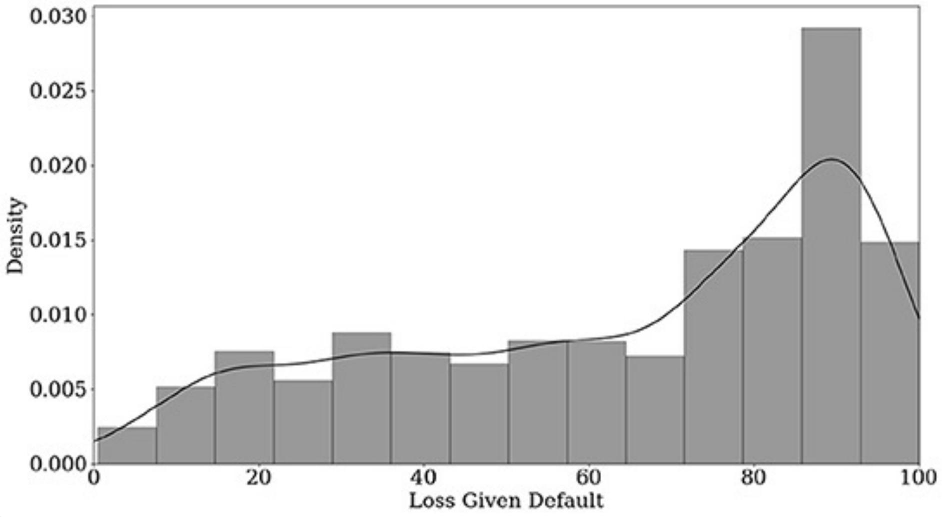
Deep Evidence Regression: Cons

- Assumes target variable (y) is Normally distributed with unknown parameters
- Not suited for Loss Given Default and other credit risk measures

$$(y_1, \ldots, y_N) \sim \mathcal{N}(\mu, \sigma^2)$$ → Not true for LGD

$$\mu \sim \mathcal{N}(\gamma, \sigma^2 \upsilon^{-1}) \qquad \sigma^2 \sim \Gamma^{-1}(\alpha, \beta).$$

Loss given defaults for Bonds



Recovery Rate (Home Loans)

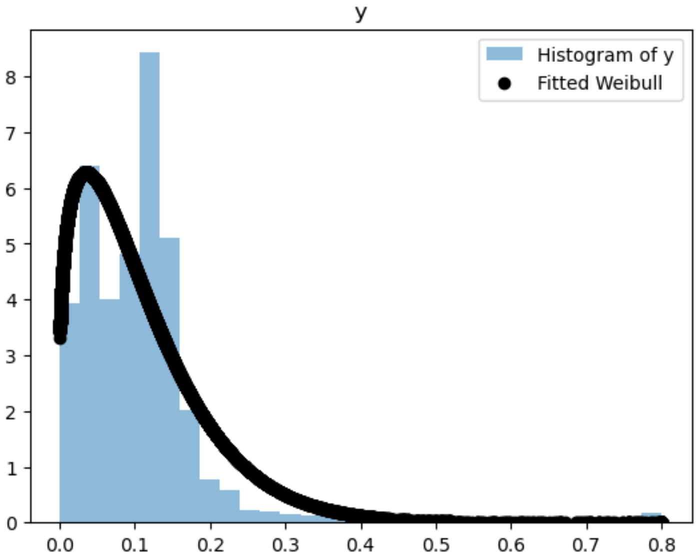


image from Max Nagl et al (2022)

Improved Deep Evidence Regression: Intuition

- What if we assume Target variable comes from Weibull Distribution
  - $y_i \sim Weibull(\lambda, k)$, k is shape parameter, λ is scale parameter
  - $f(y_i|\lambda, k) = \frac{k}{\lambda}\,(\frac{y_i}{\lambda})^{k-1}\,\exp(-\,(\frac{y_i}{\lambda})^k)$
- What are the prior distributions in that case ?
  - If we assume both k and λ are unknown
    - $k \sim \Gamma^{-1}(\alpha_1, \beta_1), \lambda \sim \Gamma^{-1}(\alpha_2, \beta_2)$
    - The computation of Log Likelihood $\mathrm{p}(\mathrm{y_i}|\alpha_1, \beta_1, \alpha_2, \beta_2)$ is intractable

Georgia Tech

Improved Deep Evidence Regression

- Instead assume:
  - k or shape is known, and λ scale is unknown
    - k can be estimated from training data (Assumption)
  - Let $\theta = \lambda^k \sim \Gamma^{-1}(\alpha, \beta)$
  - $p(y_i \mid \theta) = \frac{k}{\theta} y_i^{k-1} \exp(-\frac{{y_i}^k}{\theta})$
  - $p(y_i \mid \alpha, \beta) = \int \frac{k}{\theta} y_i^{k-1} \exp(-\frac{{y_i}^k}{\theta}) * \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{1}{\theta^{\alpha+1}} \exp(-\frac{\beta}{\theta}) \, d\theta$
  - $\boxed{p(y_i \mid \alpha, \beta) = \frac{\alpha\, k\, y_i^{k-1} \beta^\alpha}{(y_i^k + \beta)^{\alpha+1}}}$

Log Likelihood of y given evidential parameters

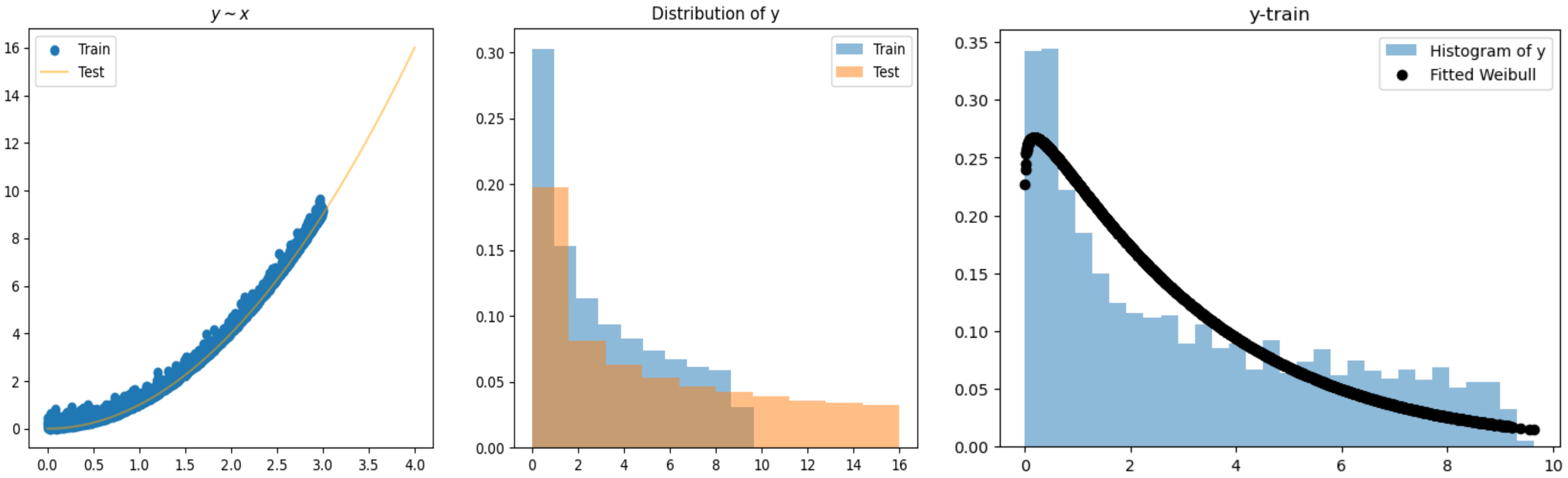
Mean Prediction

  - $\boxed{Z = E[y_i|\alpha,\beta] = \Gamma\left(1+\frac{1}{k}\right) E(\lambda|\alpha,\beta) = \Gamma\left(1+\frac{1}{k}\right) \frac{1}{\Gamma(\alpha)} \Gamma\left(\alpha - \frac{1}{k}\right) \beta^{1/k}}$

Georgia Tech

Results: Simulated Data

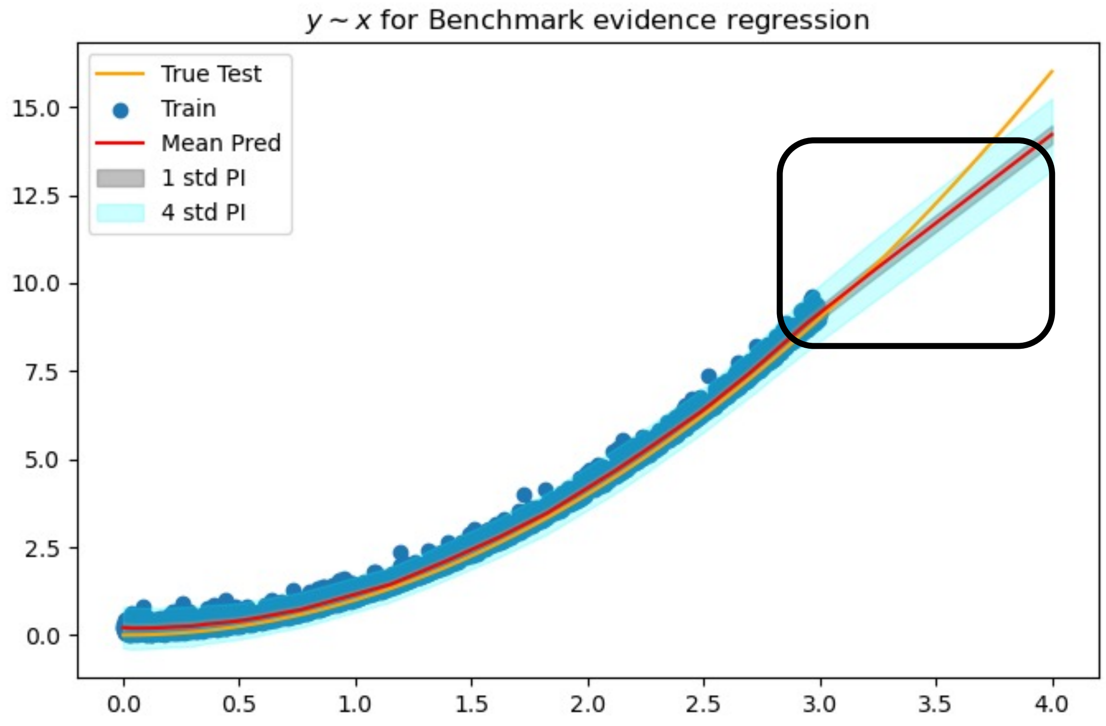
- Data generated by $y = x^2 + \epsilon, \epsilon \sim Weibull(k = 1.2, \lambda = 0.2)$
  - Train set, $x \in [0,3]$
  - Test set, $x \in [0,4]$
  - K from Weibull fit on train data = 1.05

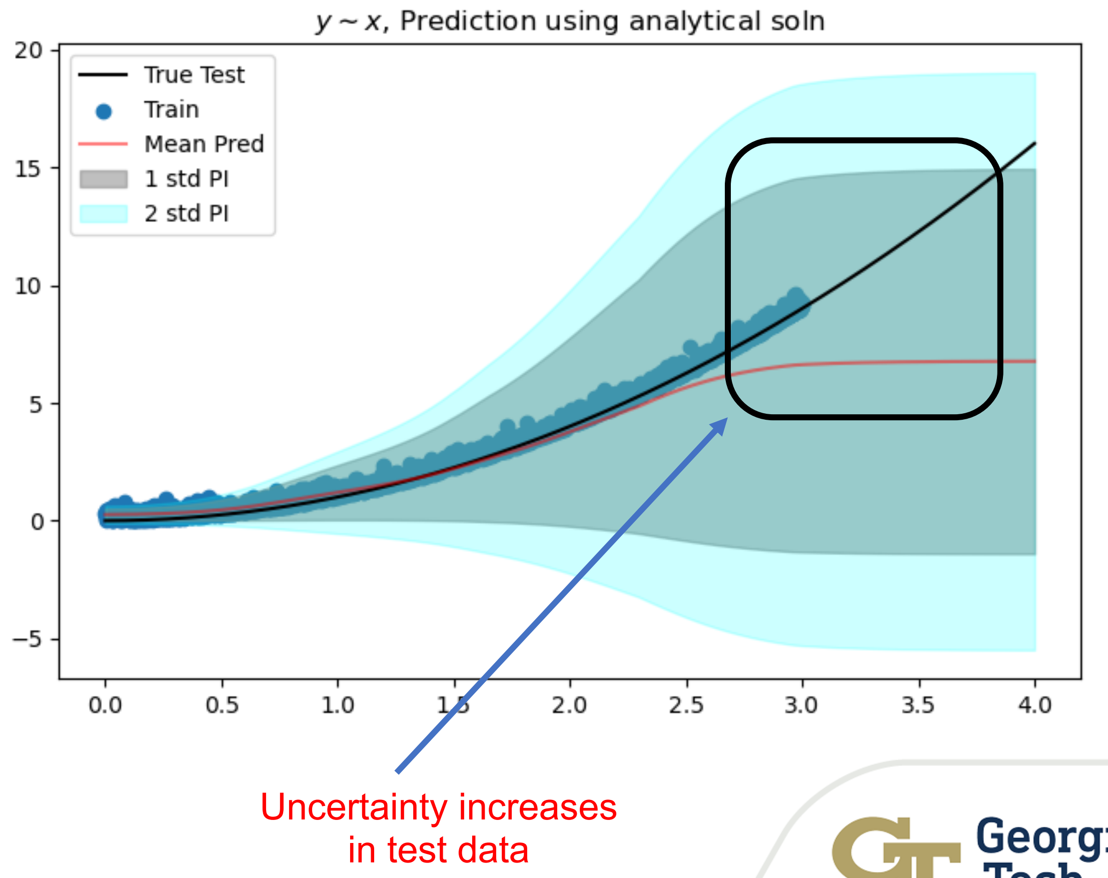


Georgia Tech

Results: Simulated Data

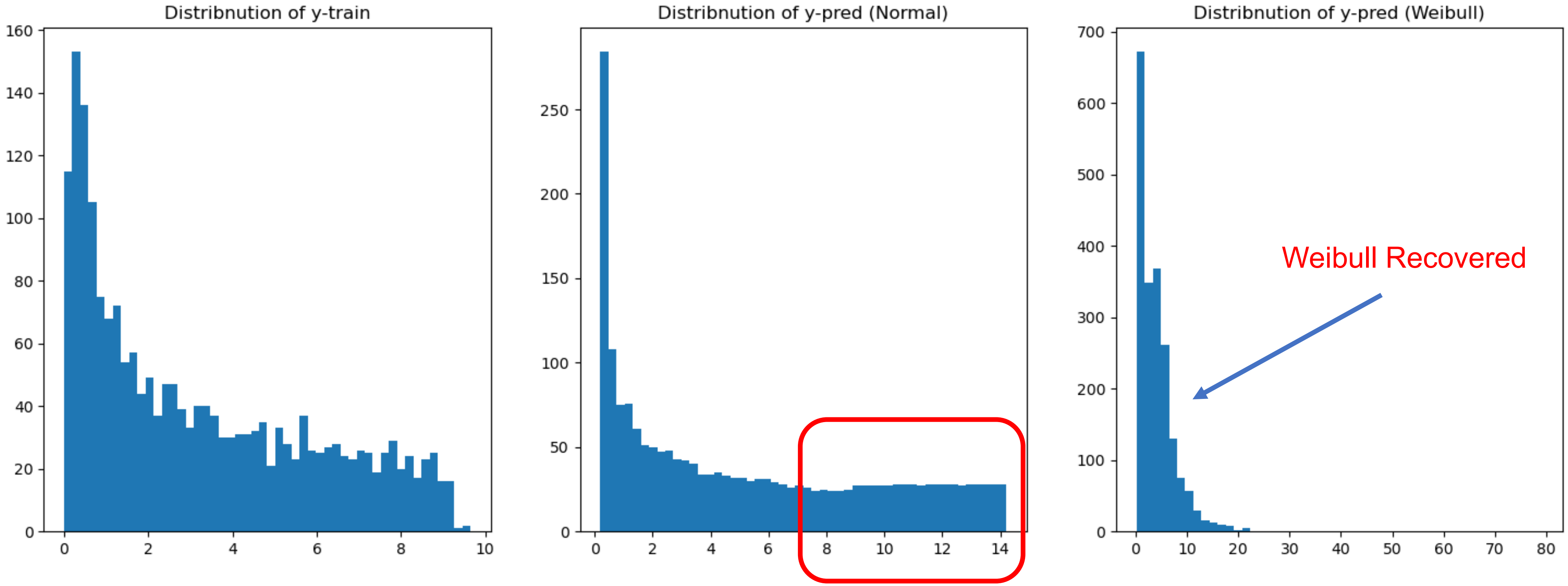
- For Weibull data Normal Evidence not able to capture Uncertainty for OOD
- Improved version captures uncertainty but not able to capture true function





Results: Simulated Data

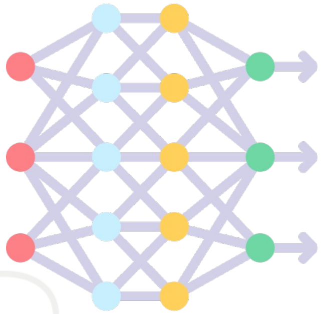
- Improved version also better at recovering original distribution of target



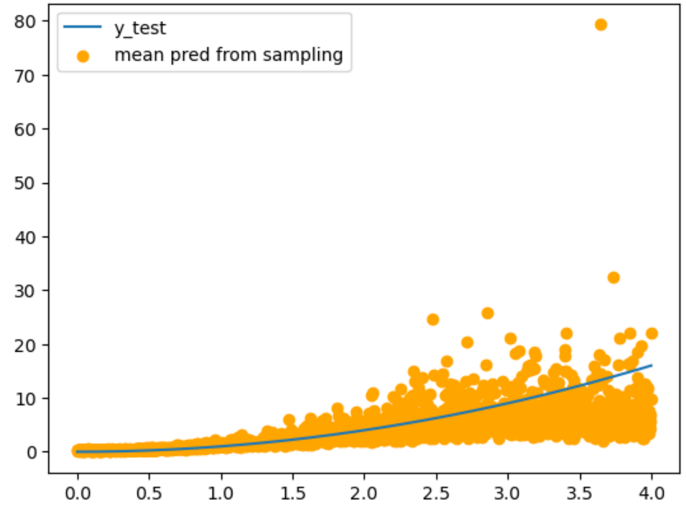
Georgia Tech

Results: Evaluation of analytical calculations

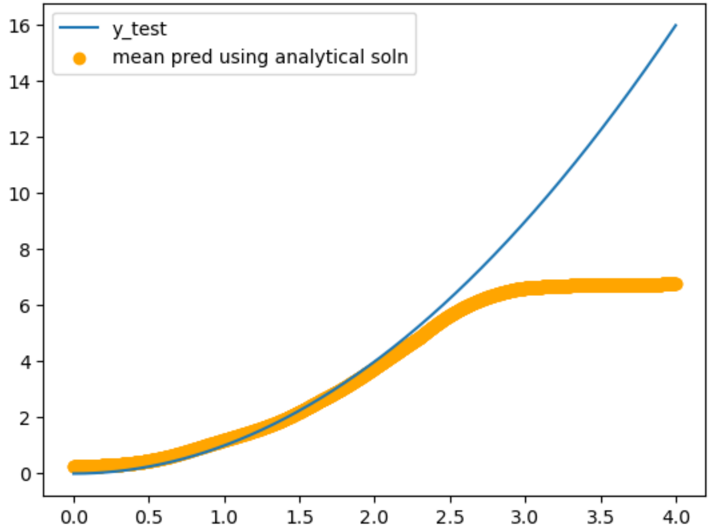
- Results from analytical computation match with those generated from sampling



$NN\ output = \{\alpha, \beta\}$



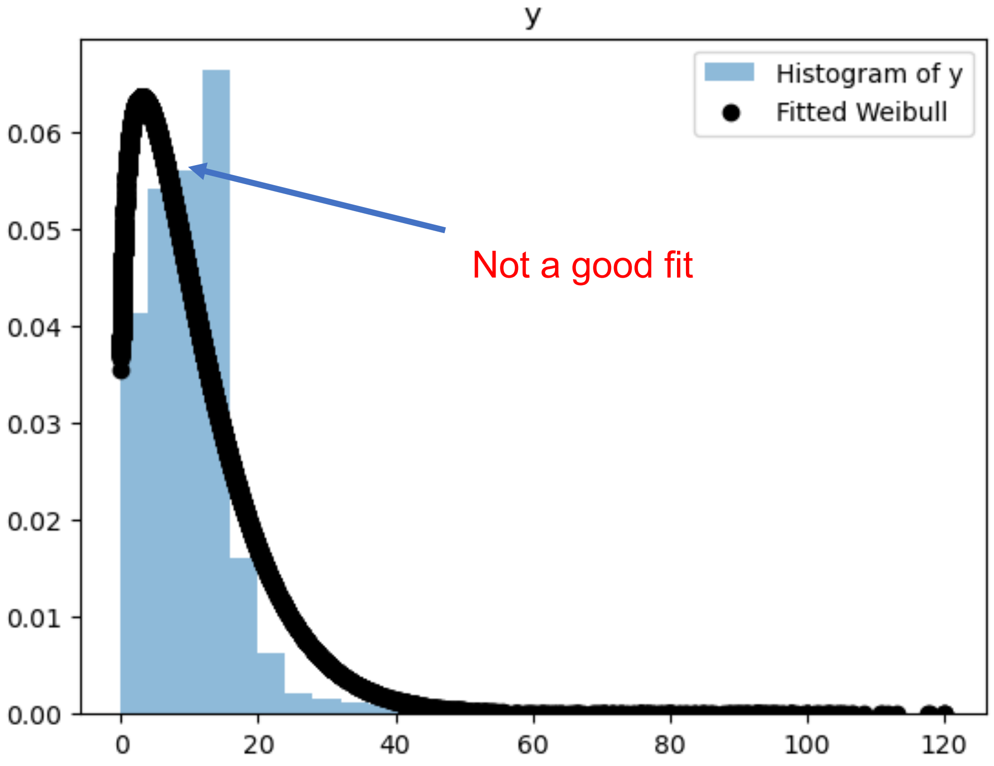
$$\theta \sim \Gamma^{-1}(\alpha, \beta)$$
$$\lambda = \theta^{1/k}$$
$$y_i \sim \mathrm{Weibull}(k, \lambda)$$

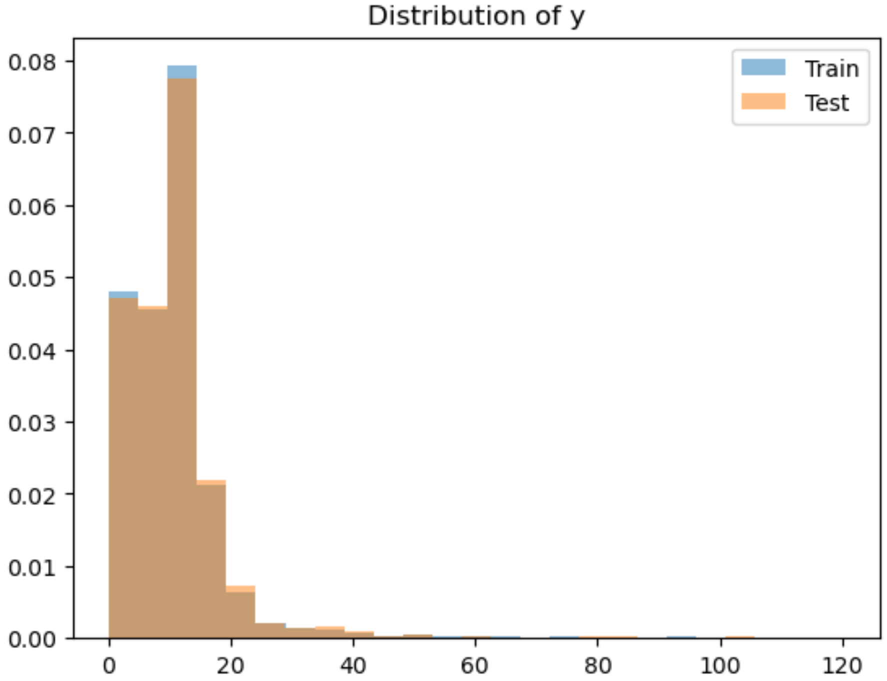


$$Z = E[y_i|\alpha, \beta] = \Gamma\left(1 + \frac{1}{k}\right) E(\lambda|\alpha, \beta) = \Gamma\left(1 + \frac{1}{k}\right) \frac{1}{\Gamma(\alpha)} \Gamma\left(\alpha - \frac{1}{k}\right) \beta^{1/k}$$

Results: Real Data

- Mortgage data from Peer to Peer lending (2007-20011)
  - Loss given default unavailable in data
  - Recovery rate used as proxy = (recoveries made/origination amount)
  - 43 'x' variables in data: time since loan, dti, joint, delinquency etc
  - ~23k rows in data

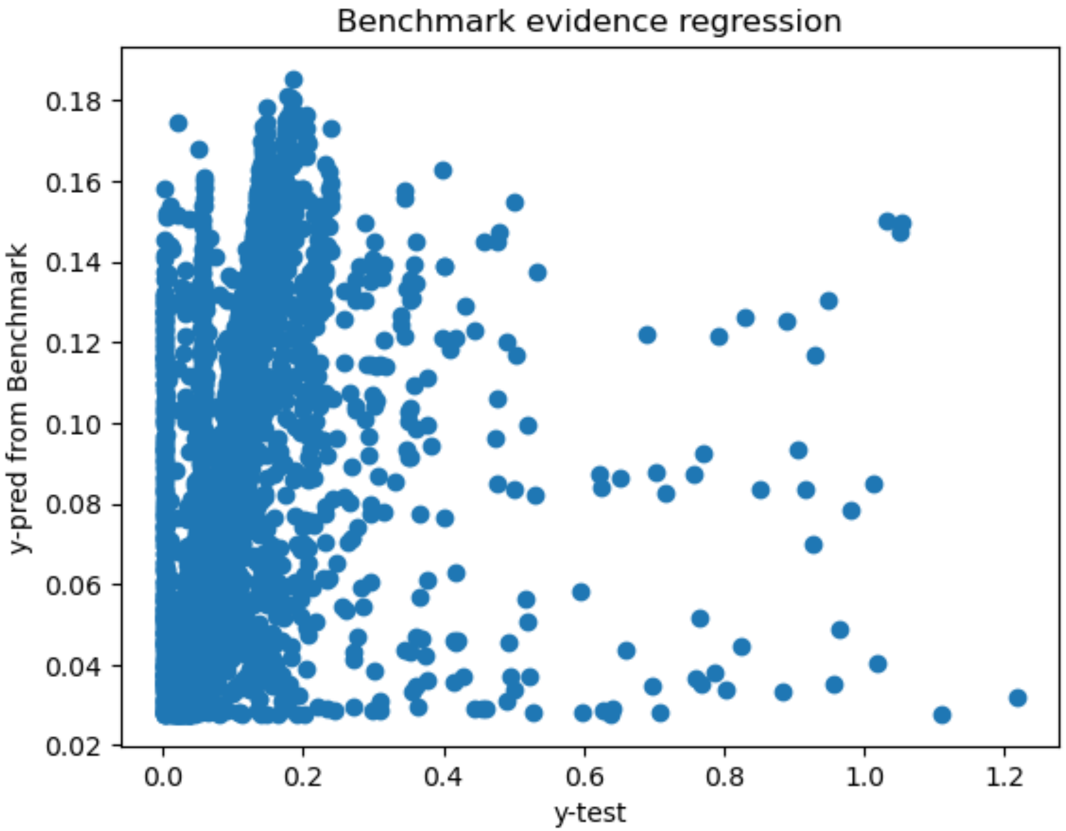


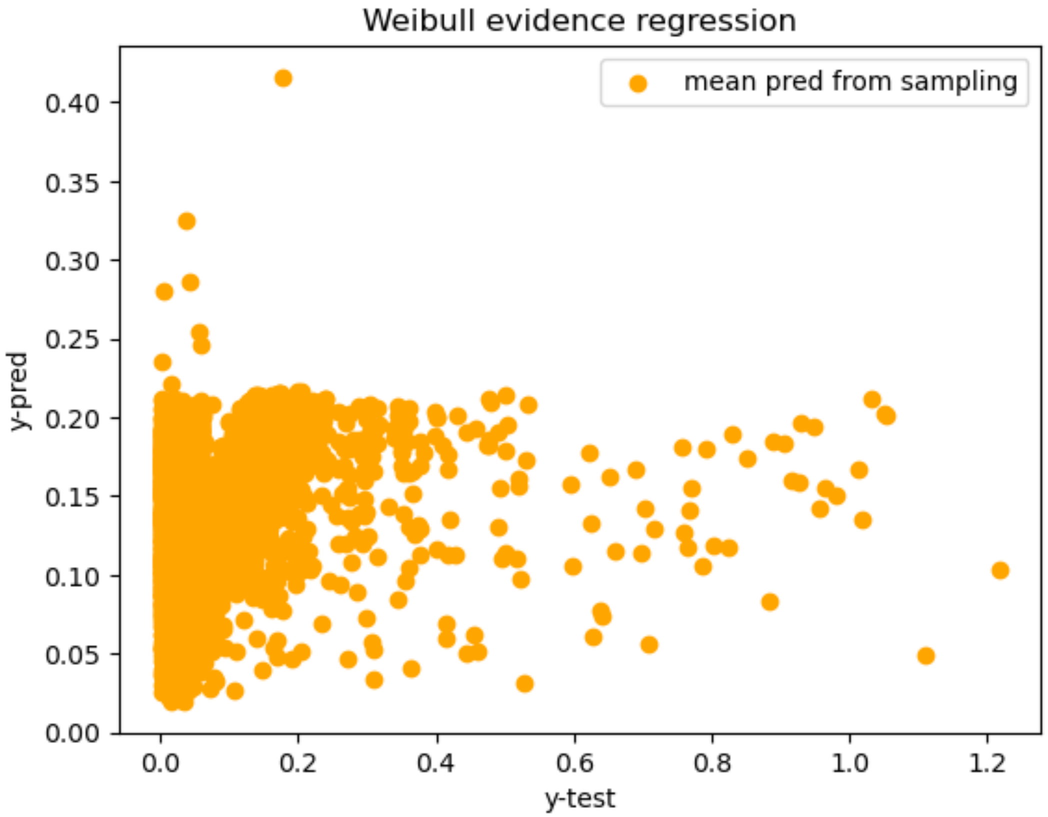


Georgia Tech

Results: Real Data

- Mortgage data from Peer to Peer lending (2007-20011)





Limitations

- Unsure of Variance calculation
- Both models very sensitive to Regularization cost
- Improved might only work for Weibull targets
- Improved requires more deep network
  - Benchmark had 4 outputs from NN – more flexible
  - Improved has 2 outputs so less flexibility

Georgia Tech

Discussion/Questions

- The presentation evolved into a very fruitful discussion, with key questions on:
  - ***How is Deep Evidential learning helpful:*** We discussed how with Evidential learning one can quantify uncertainty without sampling or keeping multiple copies of the network.
  - ***How is log-likelihood calculated:*** We summarized that log likelihood is calculated by finding joint probability of parameters of target variables given prior(evidential) parameters, and then marginalizing over the target variables parameters.
  - ***Why do we need Weibull distribution:*** We contextualized the use of Weibull distribution, showing insufficiency of normal distribution to model LGD. Furthermore, recent back collapse of SVB bank was shared as an example.
  - ***Why can we not keep both Weibull parameters unknown:*** It was noted that with both the parameters unknown, the log likelihood integral becomes intractable.
  - ***Why choose the NIG prior after reparametrizing Weibull:*** Following from the last discussion we present that this NIG prior is more amenable for analytical calculations, as the NIG lies in the conjugate family of Weibull.

Georgia Tech

Appendix: Log Likelihood Calculation

2.2. Learning Log-Likelihood. Hence we can define likelihood of y_i given the higher order evidential parameters α, β can be defined as :

$$Lik = p(y_i|\alpha,\beta) = \int_\theta p(y_i|\theta,k)p(\theta|\alpha,\beta)d\theta$$

$$\text{Now given } \lambda, k > 0 \implies \theta > 0$$

$$p(y_i|\alpha,\beta) = \int_{\theta=0}^{\infty} \big(\frac{k}{\theta}{y_i}^{k-1}\exp\left(-y_i^k/\theta\right)\big)\big(\frac{\beta^\alpha}{\Gamma(\alpha)}\frac{1}{\theta^{\alpha+1}}\exp\left(-\frac{\beta}{\theta}\right)\big)d\theta \tag{2.3}$$

$$= ky_i^{k-1}\frac{\beta^\alpha}{\Gamma(\alpha)}\int_{\theta=0}^{\infty}\frac{1}{\theta^{\alpha+2}}\exp\left(-\frac{y_i^k+\beta}{\theta}\right)$$

4

Now let

$$\Omega = ky_i^{k-1}\frac{\beta^\alpha}{\Gamma(\alpha)} \quad \text{and,}$$

$$a = y_i^k + \beta$$

$$\implies p(y_i|\alpha,\beta) = \Omega\int_{\theta=0}^{\infty}\frac{1}{\theta^{\alpha+2}}\exp\left(-\frac{a}{\theta}\right)d\theta \tag{2.4}$$

Substituting $t = 1/\theta$

$$\implies p(y_i|\alpha,\beta) = \Omega\int_{\theta=\infty}^{0} t^{\alpha+2}\exp(-at)(-dt * t^{-2})$$

$$= \Omega\int_{\theta=0}^{\infty} t^{\alpha}\exp(-at)dt$$

Now we know from [10]

$$\int_0^\infty x^n\exp(-ax)dx = \frac{\Gamma(1+n)}{a^{1+n}}$$

$$\implies p(y_i|\alpha,\beta) = \Omega\frac{\Gamma(1+\alpha)}{a^{1+\alpha}} \tag{2.5}$$

substituting back Ω and a

$$\implies Lik = p(y_i|\alpha,\beta) = ky_i^{k-1}\frac{\beta^\alpha}{\Gamma(\alpha)}\frac{\Gamma(1+\alpha)}{(y_i^k+\beta)^{1+\alpha}}$$

$$= \frac{\alpha ky_i^{k-1}\beta^\alpha}{(y_i^k+\beta)^{\alpha+1}}$$

Hence the log-likelihood for i'th observation is defined as:

$$Log-Lik_i = L_i^{lik} = \log\alpha_i + \log k + (k-1)\log y_i + \alpha_i\log\beta_i - (\alpha_i+1)({y_i}^k+\beta_i) \tag{2.6}$$

Georgia Tech

Appendix: Mean and variance calculation

2.3.1. Mean Prediction.

We define the mean prediction as

$$Z = E[y_i|\alpha,\beta]$$

Now given $y_i \sim Weibull(k,\lambda)$

$$Z = E[\lambda * \Gamma(1+\frac{1}{k})] = E(\lambda) * \Gamma(1+\frac{1}{k}) \quad \text{(k is known)} \tag{2.7}$$

$$E[\lambda] = \int_\lambda \lambda p(\lambda) d\lambda$$

Hence to solve for mean prediction we need to find pdf $p(\lambda)$. Because we know $\theta = \lambda^k \sim \Gamma^{-1}(\alpha,\beta)$, we can use change of variable to find pdf of λ [9].

$$\begin{aligned} p(\lambda|\alpha,\beta) &= p_\theta(\lambda^k) * |\frac{d\lambda^k}{d\lambda}| \\ &= \frac{\beta^\alpha}{\Gamma(\alpha)}(\frac{1}{\lambda^k})^{\alpha+1} \exp(-\frac{\beta}{\lambda^k}) * |k\lambda^{k-1}| \\ &= \frac{\beta^\alpha}{\Gamma(\alpha)}(\frac{1}{\lambda^k})^{\alpha+1} \exp(-\frac{\beta}{\lambda^k}) * k\lambda^{k-1} \quad \text{(given } \lambda, k > 0) \end{aligned} \tag{2.8}$$

Hence,

$$E[\lambda] = \int_\lambda \lambda \frac{\beta^\alpha}{\Gamma(\alpha)}(\frac{1}{\lambda^k})^{\alpha+1} \exp(-\frac{\beta}{\lambda^k}) * k\lambda^{k-1} d\lambda$$

$$= \frac{k\beta^\alpha}{\Gamma(\alpha)} \int_{\lambda=0}^{\infty} \frac{1}{\lambda^{k\alpha+k-k}} \exp(\frac{\beta}{\lambda^2}) d\lambda$$

Substituting $t = 1/\lambda$, we get:

$$dt = -1/\lambda^2 d\lambda,$$

$$E[\lambda] = \frac{k\beta^\alpha}{\Gamma(\alpha)} \int_{t=0}^{\infty} t^{k\alpha-2} \exp(-\beta t^k) dt \tag{2.9}$$

By table of integrals at [10]

$$\int_0^\infty y^m e^{-by^k}, dy = \frac{\Gamma\left(\frac{m+1}{k}\right)}{kb^{(m+1)/k}}$$

$$\implies E[\lambda] = \frac{k\beta^\alpha}{\Gamma(\alpha)} * \Gamma(\frac{k\alpha-1}{k}) * \frac{1}{k} * \frac{1}{\beta^{\frac{k\alpha-1}{k}}}$$

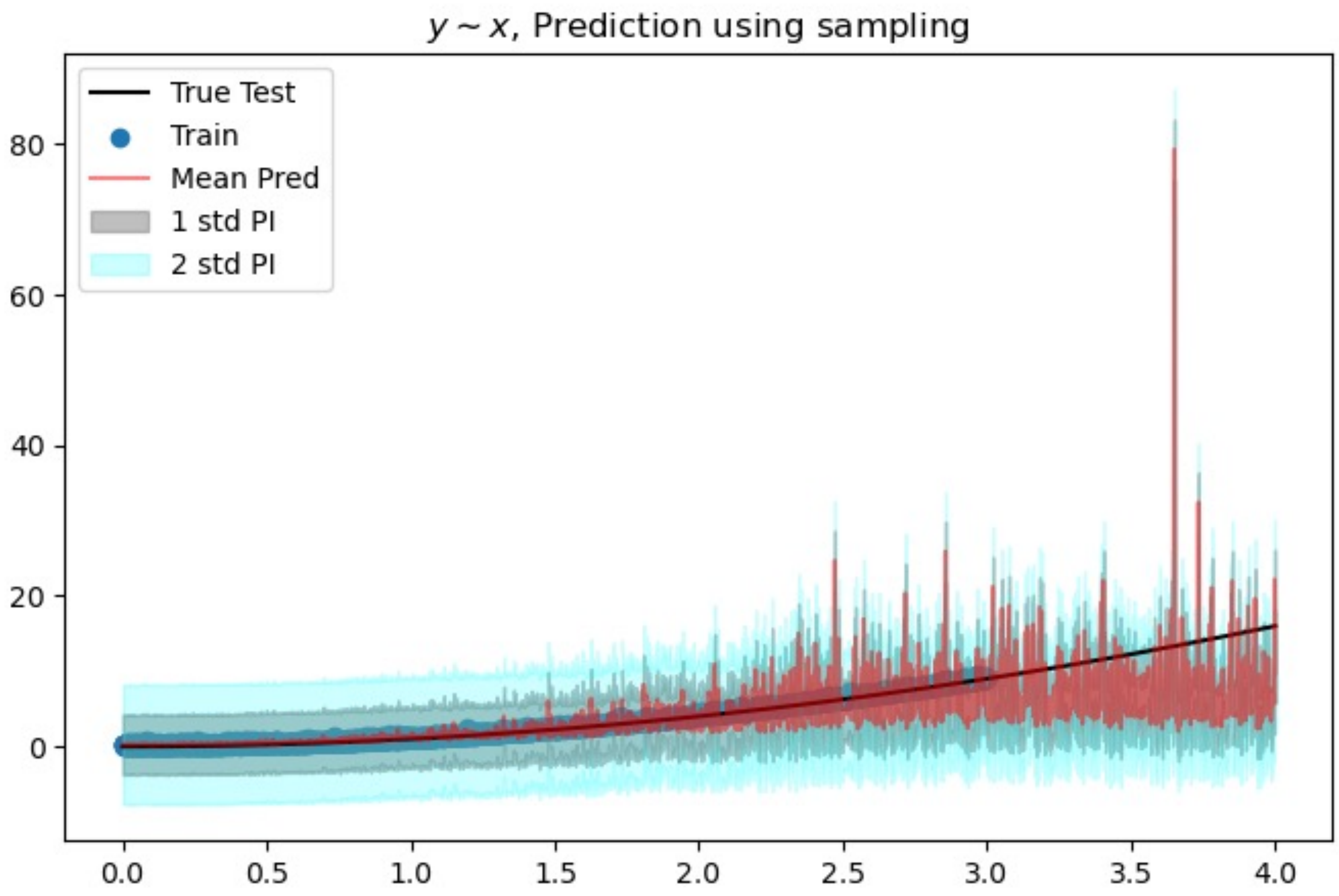
Hence we get the mean prediction as:

$$Z = E[y_i|\alpha,\beta] = E(\lambda) * \Gamma(1+\frac{1}{k})$$

$$\frac{k\beta^\alpha}{\Gamma(\alpha)} * \Gamma(\frac{k\alpha-1}{k}) * \frac{1}{k} * \frac{1}{\beta^{\frac{k\alpha-1}{k}}} * \Gamma(1+\frac{1}{k}) \tag{2.10}$$

$$= \Gamma(1+\frac{1}{k}) \frac{1}{\Gamma(\alpha)} \Gamma(\alpha - \frac{1}{k}) * \beta^{1/k}$$

Appendix: Simulated data PI from sampling



Georgia Tech

Appendix Results: Real Data

- Mortgage data from Peer to Peer lending (2007-20011)

	Relative MSE
Benchmark	16.3%
Improved	9.7%

Georgia Tech

Appendix: Simulated data PI from sampling

```
Index(['Unnamed: 0', 'id', 'member_id', 'loan_amnt', 'funded_amnt',
'funded_amnt_inv', 'term', 'int_rate', 'installment', 'grade', 'sub_grade', 'emp_title',
'emp_length', 'home_ownership', 'annual_inc', 'verification_status', 'issue_d',
'loan_status', 'pymnt_plan', 'url', 'desc', 'purpose', 'title', 'zip_code', 'addr_state', 'dti',
'delinq_2yrs', 'earliest_cr_line', 'inq_last_6mths', 'mths_since_last_delinq',
'mths_since_last_record', 'open_acc', 'pub_rec', 'revol_bal', 'revol_util', 'total_acc',
'initial_list_status', 'out_prncp', 'out_prncp_inv', 'total_pymnt', 'total_pymnt_inv',
'total_rec_prncp', 'total_rec_int', 'total_rec_late_fee', 'recoveries',
'collection_recovery_fee', 'last_pymnt_d', 'last_pymnt_amnt', 'next_pymnt_d',
'last_credit_pull_d', 'collections_12_mths_ex_med', 'mths_since_last_major_derog',
'policy_code', 'application_type', 'annual_inc_joint', 'dti_joint',
'verification_status_joint', 'acc_now_delinq', 'tot_coll_amt', 'tot_cur_bal',
'open_acc_6m', 'open_il_6m', 'open_il_12m', 'open_il_24m', 'mths_since_rcnt_il',
'total_bal_il', 'il_util', 'open_rv_12m', 'open_rv_24m', 'max_bal_bc', 'all_util',
'total_rev_hi_lim', 'inq_fi', 'total_cu_tl', 'inq_last_12m'], dtype='object')
```