The Neural Network approach to PDF fitting

Alberto Guffanti

University of Edinburgh

On behalf of the NNPDF Collaboration:

R. D. Ball, L. Del Debbio, M. Ubiali (Edinburgh), S. Forte (Milano), J. I. Latorre (Barcelona), J. Rojo-Chacon (LPTHE - Paris) and A. Piccione (Genova)

QCD@Work 2007

Martina Franca, June 16 - 20, 2007

Outline

Motivation

- PDF uncertainties will affect all areas of phenomenology at hadron colliders.
- Past experience showed that sometimes a discrepancy between theory predictions and experimental results is not a signal of "new physics", rather "old physics" we don't fully understand
  - High-E_T jets at Tevatron,
  - leptoquarks at HERA,
  - B-production at Tevatron.
- Recent updates of parton fits caused shifts in observables' predictions outside the previously quoted error bands.
- Need for faithful estimation of errors associated with parton distribution functions.

Problem
Faithful estimation of errors

- Single quantity: 1-σ error
- Multiple quantities: 1-σ contours
- Function: need an "error band" in the space of functions (*i.e.* the probability density $\mathcal{P}[f]$ in the space of functions $f(x)$)

Expectation values are Functional integrals

$$\langle \mathcal{F}[f(x)] \rangle = \int \mathcal{D}f \mathcal{F}[f(x)] \mathcal{P}[f(x)]$$

Problem
Faithful estimation of errors

- Single quantity: 1-σ error
- Multiple quantities: 1-σ contours
- Function: need an "error band" in the space of functions (*i.e.* the probability density $\mathcal{P}[f]$ in the space of functions $f(x)$)

Expectation values are Functional integrals

$$\langle \mathcal{F}[f(x)] \rangle = \int \mathcal{D}f \mathcal{F}[f(x)] \mathcal{P}[f(x)]$$

Determine an infinite-dimensional object (a function) from a finite set of data points ... mathematically ill-defined problem.

Solution
Standard Approach

- Introduce a simple functional form with enough free parameters

$$q(x, Q^2) = x^{\alpha}(1 - x)^{\beta} P(x; \lambda_1, ..., \lambda_n).$$

- Fit parameters minimizing χ^2.

Solution
Standard Approach

- Introduce a simple functional form with enough free parameters

$$q(x, Q^2) = x^{\alpha}(1-x)^{\beta} P(x; \lambda_1, ..., \lambda_n).$$

- Fit parameters minimizing χ^2.

Open problems:

- Error propagation from data to parameters and from parameters to observables is not trivial.
- Theoretical bias due to the chosen parametrization is difficult to assess.

The Strategy
Bayesian Inference Method

[Giele, Keller and Kosower, hep-ph/0104052]

- Generate a Monte-Carlo sampling of the function space according to a *reasonable* prior distribution.
- Compute observables as functional integrals with the probability measure defined by the sampling.
- Update probability using Bayesian inference on the MC sample.
- Iterate until convergence is reached.

The originally "infinite dimensional" problem is made finite by choosing a prior, but the final result should not depend on this choice.

The Neural Network Approach

1. Generate N_{rep} Monte-Carlo replicas of the experimental data.
2. *Train* a Neural Network on any of the replicas, defining a probability density on the space of the observable.
3. Expectation values for observables are sums over nets

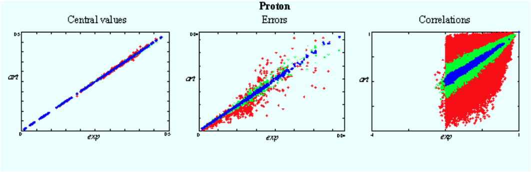
$$\langle \mathcal{F}[f(x,Q^2)] \rangle = \frac{1}{N_{rep}} \sum_{k=1}^{N_{rep}} \mathcal{F}\left(f^{(net)(k)}(x,Q^2) \right)$$

Monte Carlo replicas generation

- Generate N_{rep} Monte-Carlo replicas of the data according to

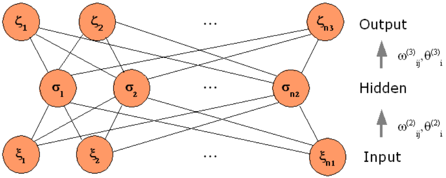
$$F_i^{(art)(k)} = (1 + r_N^{(k)}\sigma_N)\left(F_i(exp) + \sum_{p=1}^{N_{sys}} r_p^{(k)}\sigma_{i,p} + r_i^{(k)}\sigma_{i,s}\right)$$

- Validate Monte-Carlo replicas against experimental data. (statistical estimators, faithful representation of errors, convergence rate increasing N_{rep})



- $\mathcal{O}(1000)$ replicas needed to reproduce correlations to percent accuracy.

Neural Networks



- Neural Networks are a class of algorithms suitable to fit noisy or incomplete data.

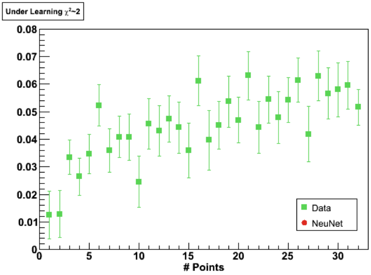
  [for HEP applications see ACAT 2007]

- Any continuous function can be approximated with neural network with one internal layer and non-linear neuron activation function.

  [K. Hornik, M. Stinchcombe and H. White (1989)]

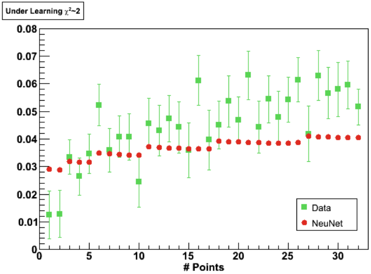
Neural Networks
Training

- Set network parameters randomly.
- If there are different inputs, normalize them.
- Define a *figure of merit* *E* (*i.e.* χ^2).
- Define a criterion of convergence (*i.e.* $\chi^2 \sim 1$).



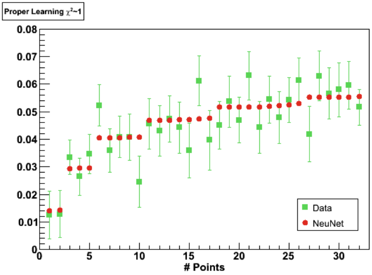
Neural Networks
Training

- Set network parameters randomly.
- If there are different inputs, normalize them.
- Define a *figure of merit* *E* (*i.e.* χ^2).
- Define a criterion of convergence (*i.e.* $\chi^2 \sim 1$).



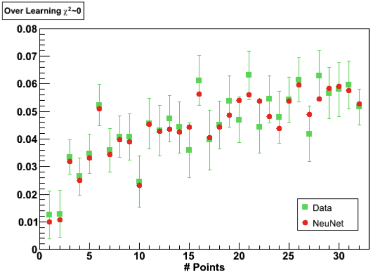
Neural Networks
Training

- Set network parameters randomly.
- If there are different inputs, normalize them.
- Define a *figure of merit* *E* (*i.e.* χ^2).
- Define a criterion of convergence (*i.e.* $\chi^2 \sim 1$).



Neural Networks
Training

- Set network parameters randomly.
- If there are different inputs, normalize them.
- Define a *figure of merit* *E* (*i.e.* χ^2).
- Define a criterion of convergence (*i.e.* $\chi^2 \sim 1$).



Neural Networks
Training Method

Which training algorithm should we use?

Neural Networks
Training Method

Which training algorithm should we use?

Genetic Algorithm

1. Set network parameters randomly.
2. Make *clones* of the set of parameters.
3. Mutate each clone.
4. Evaluate χ^2 for all the clones.
5. Select the clone that has the lowest χ^2.
6. Back to 2, until stability in χ^2 is reached.

Neural Networks
Training Method

Which training algorithm should we use?

Genetic Algorithm

1. Set network parameters randomly.
2. Make *clones* of the set of parameters.
3. Mutate each clone.
4. Evaluate χ^2 for all the clones.
5. Select the clone that has the lowest χ^2.
6. Back to 2, until stability in χ^2 is reached.

Pros:

- Allows to minimize the fully correlated χ^2.
- Explores the full parameter space, reducing the risk of being trapped in a local minimum.

Cons:

- Slow convergence.
- χ^2 decreases monotonically - need to find a suitable stopping criterion.

Neural Networks
Stopping criterion

When to stop a fit to avoid overlearning?

Neural Networks
Stopping criterion

When to stop a fit to avoid overlearning?

Stopping criterion based on Training-Validation separation

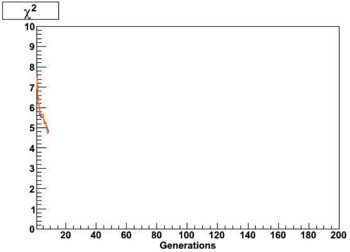
- Divide the data in two sets: Training and Validation.
- Minimize the χ^2 of the data in the Training set.
- Compute the χ^2 for the data in the Validation set.
- When Validation χ^2 stops decreasing, **STOP** the fit.

Neural Networks
Stopping criterion

When to stop a fit to avoid overlearning?

Stopping criterion based on Training-Validation separation

- Divide the data in two sets: Training and Validation.
- Minimize the χ^2 of the data in the Training set.
- Compute the χ^2 for the data in the Validation set.
- When Validation χ^2 stops decreasing, **STOP** the fit.

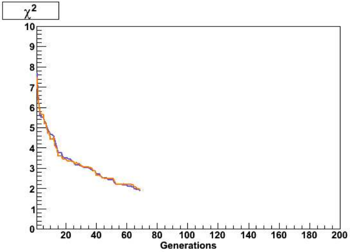


Neural Networks
Stopping criterion

When to stop a fit to avoid overlearning?

Stopping criterion based on Training-Validation separation

- Divide the data in two sets: Training and Validation.
- Minimize the χ^2 of the data in the Training set.
- Compute the χ^2 for the data in the Validation set.
- When Validation χ^2 stops decreasing, **STOP** the fit.

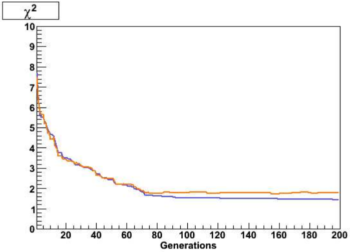


Neural Networks
Stopping criterion

When to stop a fit to avoid overlearning?

Stopping criterion based on Training-Validation separation

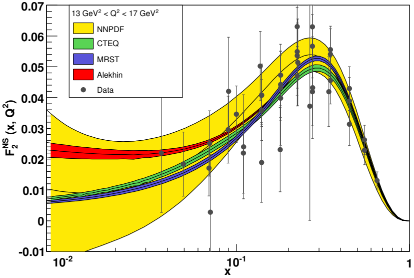
- Divide the data in two sets: Training and Validation.
- Minimize the χ^2 of the data in the Training set.
- Compute the χ^2 for the data in the Validation set.
- When Validation χ^2 stops decreasing, **STOP** the fit.



Non Singlet Analysis

F_2^{NS} determination

[L. Del Debbio et al., hep-ph/0701127]



- Compatible with results from other PDF determinations (even when they are not in agreement)
- Lager uncertainties both in the
  - Data region (MC error estimation)
  - Extrapolation region (functional form bias)

NNPDF - The full set
Status report

- Increased complexity related to:
  - Full DGLAP evolution
  - Training multiple Neural Networks at the same time
- First preliminary fits run smoothly providing a proof-of-concept of the feasibility of the whole project

Summary
Where we are ...

- Standard approaches to PDF fitting might lead to underestimation of errors associated with parton densities.
- Combination of Monte-Carlo sampling techniques and Neural Networks as unbiased interpolation functions recently proved to be a reliable alternative.
- The first results concerning the determination of the quark isotriplet parton distribution have been published.

Instead of Conclusions

The way ahead ...

- The extension of the results to a full global PDF fit is at an advanced stage.
- All major technical issues have been tackled and the first preliminary results look encouraging.
- First full NNPDF fit to be expected in Autumn 2007.