[Paper review 12]

Representing Inferential Uncertainty in Deep Neural Networks through Sampling

(Patrick McClure & Nickolaus Kriegeskorte , 2017)

[Contents]

0. Abstract

Bayesian models catches model uncertainty

(recent work : dropout-based variational distribution)

In this paper, evaluate Bayesian DNN trained with

- 1) Bernoulli drop out
- 2) Bernoulli drop connect
- 3) Gaussian drop out
- 4) Gaussian drop connect
- 5) (new) spike-and-slab

1. Introduction

BNN learns "distribution over parameters" $\rightarrow$ offer "uncertainty estimates"

However, these do not scale well! (difficulty in computing posterior)

How to find posterior? Example :

1) HMC (Hamiltonian Monte Carlo) (Neal, 2012)

- use the gradient information calculated using back-prop to perform MCMC

2) Approximate method

- Variational inference,

3) Dropout, Drop=connect...

In this paper, "investiage how using MC sampling to model uncertainty affects a network's probabilistic predictions"

Use variational distributions, based on 1)~5) (in 0.Abstract)

2. Methods

2.1 BNN

- using VI, $q(W)$ is learned by maximizing ELBO
  (= minimizing : $-\int \log p\left(D_{train} \mid W\right) q(W) dW + KL(q(W) \| p(W))$)
- to estimate the probability of test data, using $q(W) \rightarrow$ use MC sampling
  $p\left(D_{\text{test}}\right) \approx \frac{1}{n} \sum_{i}^{n} p\left(D_{\text{test}} \mid \hat{W}^{i}\right) \text{ where } \hat{W}^{i} \sim q(W)$

2.2 Variational Distributions

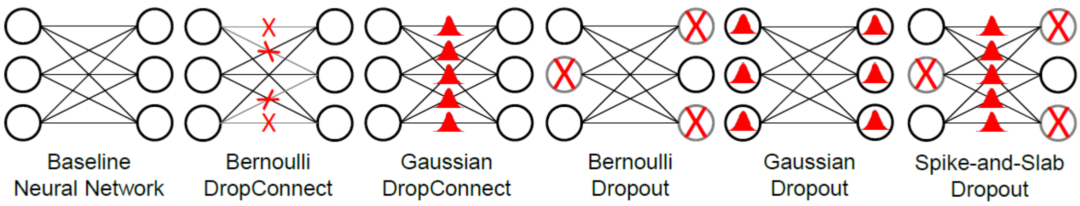
number of parameters in DNN $\rightarrow$ computationally challenging

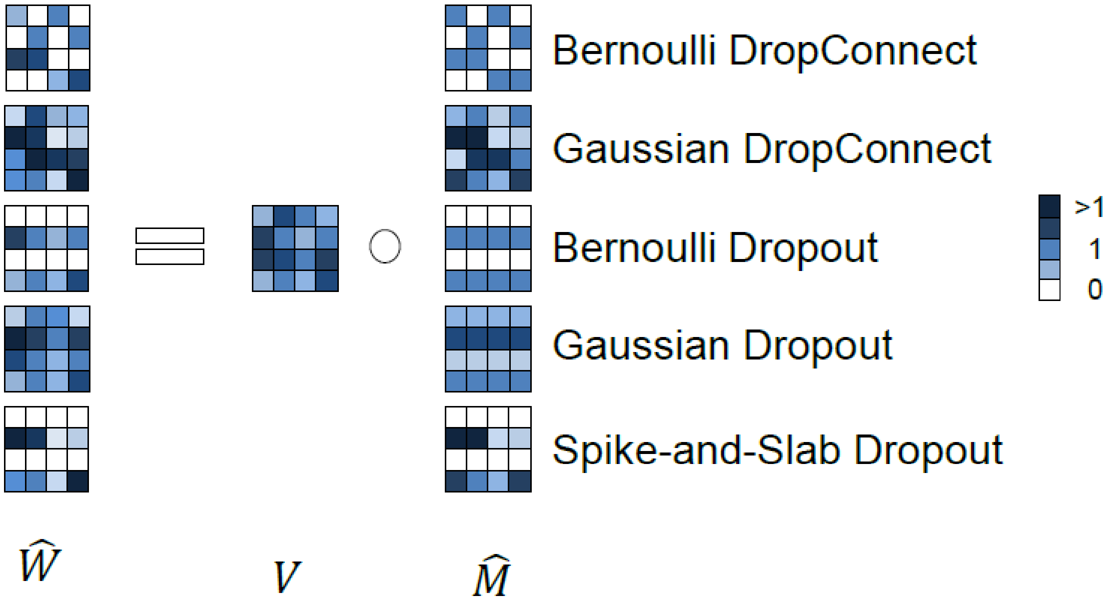
use "Variational Distribution" to sample easily!

- ex) dropout, drop connect...

$\hat{W}=V \circ \hat{M} \text{ where } \hat{M} \sim p(M)$

- $\hat{M}$: mask
- V : variational parameters
- (difference of dropout & drop connect : just the probability distribution used to generate the Mask !)





2.2.1 "Bernoulli" drop out & drop connect

Drop-out

- $\hat{m}_{i,*}$ ~Bernoulli(p)
- just a special case of drop-connect (all j's are same)

Drop-connect

- $\hat{m}_{i,j}$ ~Bernoulli(p)

2.2.2 "Gaussian" drop out & drop connect

(Srivastava et al, 2014) proposed Gaussian distribution with

- mean : 1
- variance : $\sigma_{dc}^2 = (1-p)/p,$

Drop-out

- $\hat{m}_{i,*} \sim \mathcal{N}\left(1, \sigma_{dc}^2\right)$
- just a special case of drop-connect (all j's are same)

Drop-connect

- $\hat{m}_{i,j} \sim \mathcal{N}\left(1, \sigma_{dc}^2\right).$

2.2.3 Spike-and-Slab Dropout

Spike-and-Slab distribution

- normalized linear combination of "spike" (of a probability mass at zero)

  and "slab" consisting of Gaussian distribution
- With probability
  - p_{spike} : return 0
  - $1-p_{\text{spike}}$: random sample from $\mathcal{N}\left(\mu_{\text{slab}}, \sigma_{\text{slab}}^2\right)$.

Use "Bernoulli dropout & Gaussian drop connect" to approximate Spike-and-Slab distribution

(by optimizing lower-bound of objective function)

- $m_{i,j} \sim b_{i,*} \mathcal{N}\left(1, \sigma_{dc}^2\right)$ where
  - $b_{i,*} \sim \operatorname{Bern}(p_{\text{do}})$
  - $\sigma_{dc}^2 = p_{dc} / \left(1-p_{dc}\right)$