There are many possible criteria for complexity estimation used in machine learning. Typical examples are Akaike's information criterion (AIC) [1] or the Bayesian Information Criterion (BIC) [63]. Their expressions are usually based on the residual sum of squares (Res) of the considered model (first term of the criterion) plus a penalty term (second term of the criterion). Differences between criteria mostly occur on this penalty term. AIC, Equation 3.20, penalizes only by the number of parameters p of the model, so that not too many free parameters are used to obtain a good fit by the model. On the other hand, BIC, Equation 3.21, takes into account also the number of samples N used for the model training.

$$AIC = N \times \log\left(\frac{Res}{N}\right) + 2 \times p. \tag{3.20}$$

$$BIC = N \times \log\left(\frac{Res}{N}\right) + p \times \log N \tag{3.21}$$

The AIC is known to have consistency problems: the AIC, it is not guaranteed that the complexity selection [illegible] verge toward an optima if the number of samples goes [illegible]. The main idea raised by this observation is about trying to balance the underfitting and the overfitting when using such criteria. This is achieved through the penalty term, for example, by having a $\log N$ based term in the penalty (where N is the number of samples), which the BIC has.

The Hannan-Quinn Information Criterion (HQ) [38] is defined as

$$HQ = N \times \log\left(\frac{Res}{N}\right) + 2 \times p \times \log\log N. \tag{3.22}$$

The HQ is very close to the other two presented criteria, as can be seen comparing the expressions of the AIC and BIC, Equations 3.20 and 3.21, with the definition of HQ, Equation 3.22.

The idea behind the design of the HQ criterion is to provide a consistent criterion (regarding for example AIC which is not consistent in its standard definition) in which the second term (the penalty) $2 \times p \times \log\log N$ grows but at a very slow rate, regarding the number

of samples. Therefore, the HQ can be considered as a more consistent compromise between the AIC and the BIC.

Examples of the HQ criterion usage can be found from Publication

There are many possible criteria for complexity estimation used in machine learning. Typical examples are Akaike's information criterion (AIC) [1] or the Bayesian Information Criterion (BIC) [63]. Their expressions are usually based on the residual sum of squares (Res) of the considered model (first term of the criterion) plus a penalty term (second term of the criterion). Differences between criteria mostly occur on this penalty term. AIC, Equation 3.20, penalizes only by the number of parameters p of the model, so that not too many free parameters are used to obtain a good fit by the model. On the other hand, BIC, Equation 3.21, takes into account also the number of samples N used for the model training.

$$AIC = N \times \log\left(\frac{Res}{N}\right) + 2 \times p. \tag{3.20}$$

$$BIC = N \times \log\left(\frac{Res}{N}\right) + p \times \log N \tag{3.21}$$

The AIC is known to have consistency problems: while minimizing the AIC, it is not guaranteed that the complexity selection will converge toward an optima if the number of samples goes to infinity [13]. The main idea raised by this observation is about trying to balance the underfitting and the overfitting when using such criteria. This is achieved through the penalty term, for example, by having a $\log N$ based term in the penalty (where N is the number of samples), which the BIC has.

The Hannan-Quinn Information Criterion (HQ) [38] is defined as

$$HQ = N \times \log\left(\frac{Res}{N}\right) + 2 \times p \times \log\log N. \tag{3.22}$$

The HQ is very close to the other two presented criteria, as can be seen comparing the expressions of the AIC and BIC, Equations 3.20 and 3.21, with the definition of HQ, Equation 3.22.

The idea behind the design of the HQ criterion is to provide a consistent criterion (regarding for example AIC which is not consistent in its standard definition) in which the second term (the penalty) $2 \times p \times \log\log N$ grows but at a very slow rate, regarding the number

of samples. Therefore, the HQ can be considered as a more consistent compromise between the AIC and the BIC.

Examples of the HQ criterion usage can be found from Publication 7 and from [55].

There are many possible criteria for complexity estimation used in machine learning. Typical examples are Akaike's information criterion (AIC) [1] or the Bayesian Information Criterion (BIC) [63]. Their expressions are usually based on the residual sum of squares (Res) of the considered model (first term of the criterion) plus a penalty term (second term of the criterion). Differences between criteria mostly occur on this penalty term. AIC, Equation 3.20, penalizes only by the number of parameters p of the model, so that not too many free parameters are used to obtain a good fit by the model. On the other hand, BIC, Equation 3.21, takes into account also the number of samples N used for the model training.

$$AIC = N \times \log\left(\frac{Res}{N}\right) + 2 \times p. \tag{3.20}$$

$$BIC = N \times \log\left(\frac{Res}{N}\right) + p \times \log N, \tag{3.21}$$

The AIC is known to have consistency problems: while minimizing the AIC, it is not guaranteed that the complexity selection will converge toward an optima if the number of samples goes to infinity [13]. The main idea raised by this observation is about trying to balance the underfitting and the overfitting when using such criteria. This is achieved through the penalty term, for example, by having a $\log N$ based term in the penalty (where N is the number of samples), which the BIC has.

The Hannan-Quinn Information Criterion (HQ) [38] is defined as

$$HQ = N \times \log\left(\frac{Res}{N}\right) + 2 \times p \times \log\log N. \tag{3.22}$$

The HQ is very close to the other two presented criteria, as can be seen comparing the expressions of the AIC and BIC, Equations 3.20 and 3.21, with the definition of HQ, Equation 3.22.

The idea behind the design of the HQ criterion is to provide a consistent criterion (regarding for example AIC which is not consistent in its standard definition) in which the second term (the penalty) $2 \times p \times \log\log N$ grows but at a very slow rate, regarding the number

of samples. Therefore, the HQ can be considered as a more consistent compromise between the AIC and the BIC.

Examples of the HQ criterion usage can be found from Publication 7 and from [55].